Washington Assessment of Student Learning

Grade 10

1999

Technical Report

Prepared by
Catherine S. Taylor
University of Washington

for
Office of the Superintendent of Public Instruction
P.O. Box 47220
Olympia, Washington 98504-7220

April 17, 2001

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont.)

TABLE OF TABLES

TABLE OF TABLES (Cont.)

TABLE OF TABLES (Cont.)

TABLE OF FIGURES

The Washington Assessment of Student Learning: March 2000

PURPOSE OF TECHNICAL REPORT

Standards for Educational and Psychological Testing (AERA/APA/NCME, 1999) require that test developers and publishers produce a technical manual. The technical manual must provide overall information documenting the technical quality of the assessment, including evidence for the reliability and validity of test scores. This document contains the technical information for the 1999 *Washington Assessment of Student Learning*: Grade 10 Assessment for Reading, Mathematics, Listening and Writing.

PART 1

OVERVIEW

BACKGROUND FOR THE STATE ASSESSMENT PROGRAM

In 1993, Washington State embarked on the development of a comprehensive school change effort that has as its primary goal the improvement of teaching and learning. Created by the state legislature in 1993, the Commission on Student Learning was charged with three important tasks in support of this school change effort:

- to establish Essential Academic Learning Requirements (EALRs) that describe what all students should know and be able to do in eight content areas--reading, writing, communication, mathematics, science, health/fitness, social studies, and the arts;
- to develop an assessment system to measure student progress at three grade levels towards achieving the EALRs; and
- to recommend an accountability system that recognizes and rewards successful schools and provides support and assistance to less successful schools.

The Commission has achieved its first major task. The EALRs in Reading, Writing, Communications, and Mathematics were first adopted in 1995 and revised in 1997 (See Appendix A). Performance "benchmarks" were also established at three grade levels--elementary (Grade 4), middle (Grade 7), and high school (Grade 10). The EALRs for Science, Social Studies, Health/Fitness, and the Arts were initially adopted in 1996 and also revised in 1997. Performance "benchmarks" for science were also established at three grade levels--elementary (Grade 5), middle (Grade 8), and high school (Grade 10).

The Commission's second major task is to develop an assessment system to determine the extent to which students are achieving the knowledge and skills defined by the EALRs. The assessments for Reading, Writing, Communication, and Mathematics have been developed at Grades 4 and 7 and were both operational as of spring, 1998. The Grade 10 assessment in these same content areas was pilot-tested in spring, 1998 and was operational beginning spring, 1999. Participation in the Grade 4 assessment was mandatory for all public schools beginning spring, 1998. Participation in the Grade 7 and 10 assessments is voluntary until spring, 2001.

Preliminary work is underway to develop middle and high school assessments in Science beginning with pilot assessments in spring, 1998 and operational assessments in spring, 2001.

Assessment development work in the other content areas--Social Studies, Health and Fitness, and the Arts--awaits legislative approval and funding.

WASHINGTON ASSESSMENT SYSTEM

The assessment system has four major components: state-level assessments, classroom-based assessments, professional staff development, and school and system context indicators. These components are described briefly below. Two additional features, the Certificate of Mastery and the Accountability System, are also briefly described.

State-Level Assessments in Reading, Writing, Listening, and Mathematics

The state-level assessments require students to both select and create answers to demonstrate their knowledge, skills, and understanding in each of the EALRs—from multiple-choice and short-answer items to more extended responses, essays, and problem solving tasks. Student, school, and district scores are reported for the operational assessments. The state-level operational test forms are standardized and "on demand" meaning all students respond to the same items, under the same conditions, and at the same time during the school year.

All of the state-level assessments are untimed; that is, students may have as much time as they reasonably need to complete their work. Guidelines for providing accommodations to students with special needs have been developed to encourage the inclusion of as many students as possible. Special need students include those in special education programs, those with Section 504 plans, English language learners (ESL/bilingual), migrant students, and highly capable students. A broad range of accommodations allows nearly all students access to some or all parts of the assessment (see *Guidelines for Inclusion and Accommodations for Special Populations on State-Level Assessments*).

Classroom teachers and curriculum specialists from across Washington were selected to assist with the development of the items for the state-level assessments. Two content committees were created at each grade level--one for Reading/Writing/Communication and one for Mathematics. Working with content and assessment specialists from the Riverside Publishing Company (one of the Commission's assessment development contractors), these committees defined the test and item specifications consistent with the Washington State Essential Academic Learning Requirements, reviewed all items prior to pilot testing, and provided final review and approval of all items after pilot testing. A separate "fairness" committee, composed of individuals reflective of Washington's diversity, also reviewed all items for words or content that might be offensive to students or parents, or might disadvantage some students for reasons unrelated to the skill or concept being assessed. (See Part 2 for a more detailed description of this process.)

Literally hundreds of items were developed and pilot-tested to create a "pool" of items. This will allow the creation of new forms of the assessment each year by sampling from the pool. Statistical "equating" procedures are used to maintain the same performance standard from year to year and to provide longitudinal comparisons across years even though different items are used.

The state-level assessments in Reading, Communication, and Mathematics include a mix of multiple-choice, short-answer, and extended-response items. Having a large pool of items provides the opportunity to vary the kinds of items from year to year so that a particular item format (e.g. multiple-choice, short-answer, or extended-response) is not always associated with the same Essential Academic Learning Requirements. (See Part 2 for more detail on the item types)

Following the first operational assessment at each grade level, a standard-setting committee determined the level of performance on the assessments that would be required for students to "meet the standard" on the Essential Academic Learning Requirements. In addition, "progress categories" above and below the standard were established in Reading and Mathematics to show growth over time as well as to give students and parents an indication of how far from the standard in these content areas a student's performance is. School and district performance on the assessments is reported in terms of the percentage of students meeting the standard and in each of the progress categories. (See Part 5 for a complete description of the standard setting process).

An *Example Test* and *Assessment Sampler* for each of the Grade 4, 7, and 10 operational assessments were created for teachers, students, and parents. The *Example Tests* along with the *Assessment Samplers* include samples of the test items, the scoring criteria for the items, and examples of student responses that have been scored. In addition to these materials, an interactive CD-ROM system called NCS Mentor for Washington provides teachers and students with another means to review the Essential Academic Learning Requirements and practice scoring student responses to items like those contained on the operational assessments.

Classroom-Based Assessment

There were a number of important reasons for including classroom-based assessment as part of the new assessment system. First, classroom-based assessments help students and their teachers better understand the Essential Academic Learning Requirements and to recognize the characteristics of quality work that define good performance for each content area. Second, classroom-based assessments provide assessment of some of the EALRs for which state-level assessment is not feasible (for example, oral presentations or group discussion). Third, classroom-based assessments offer teachers and students opportunities to gather evidence of student achievement in ways that best fit the needs and interests of individual students. Fourth, classroom-based assessments help teachers become more effective in gathering valid evidence of student learning related to the Essential Academic Learning Requirements. And finally, good classroom-based assessments can be more sensitive to the developmental needs of students and provide the flexibility necessary to better accommodate the learning styles of children with special needs. In addition to the items that may be on the state-level assessments, classroom-based assessments can provide information from oral interviews and presentations, work products, experiments and projects, or exhibitions of student work collected over a week, a month, or the entire school year.

Classroom-based assessment *Tool Kits* have been developed for the early and middle years to provide teachers with examples of good assessment strategies. The *Tool Kits* include models for paper and pencil tasks, generic checklists of skills and traits, observation assessment strategies, simple rating scales, and generic protocols for oral communications and personal interviews. At the upper grades, classroom-based assessment strategies will also include models for developing and evaluating cross-discipline, performance-based tasks. In addition to the models, the *Tool Kits* also provide content frameworks to assist teachers, at all grade levels, to relate their classroom learning goals and instruction to the Essential Academic Learning Requirements.

Professional Development

A third major component of the new assessment system emphasizes the need for ongoing, comprehensive support and professional training for teachers and administrators to improve their understanding of the Essential Academic Learning Requirements, the characteristics of sound assessments, and effective instructional strategies that will help students reach the standards. The Commission on Student Learning established fifteen "Learning and Assessment Centers" across the

state. Most are managed through Washington's nine Educational Service Districts with a few managed by school district consortia. These Centers provide professional development and support to assist school and district staff in:

1 linking teaching and curriculum to high academic standards based on the EALRs;

2 learning and applying the principles of good assessment practice;

3 using a variety of assessment techniques and strategies;

4 judging student work by applying explicit scoring criteria;

5 making instructional and curricular decisions based on reliable and valid assessment information; and

6 helping students and parents to understand the EALRs and how students can achieve them.

Context Indicators

Context indicators help teachers, parents, and the public understand and interpret student performance in relation to the environment in which teaching and learning occur. Examples of potentially useful indicators include information about faculty experience and training, instructional strategies employed, special programs for students, condition of facilities and equipment, availability of appropriate instructional materials and technology, relevant characteristics of students and the community, student attendance patterns, grade to grade transition successes, and high school dropout and graduation rates. The purpose for context information is not to explain away or excuse low performance. Rather, context indicators can provide important information to schools, policy-makers, and the public about the conditions that support or inhibit success in helping all students achieve the Essential Academic Learning Requirements.

Certificate of Mastery

Once the Essential Academic Learning Requirements and new standards are fully in place, graduating seniors will be required to earn a Certificate of Mastery to get a high school diploma. The Certificate will serve as evidence that students have achieved Washington's Essential Academic Learning Requirements by meeting the standards set for the Grade 10 assessments. Preliminary recommendations for implementing the Certificate have been forwarded to the legislature and include the recommendation that initial use should be based only on meeting the standards in Reading, Writing, Communication, Mathematics, and Science. The Certificate as a high school graduation requirement would begin with the graduating class of 2008. The Commission recommended that meeting the standards in the other content areas be treated as "endorsements" rather than as requirements once those assessments are developed and operational.

School and District Accountability System

The Academic Achievement and Accountability (A+) Commission has developed preliminary recommendations for a school and district accountability system that will recognize schools who are successful in helping their students achieve the standards on the WASL assessments. Recommendations also address the need for assistance to those schools and districts in which students are not achieving the standards. The task force recommendations are currently in draft form and are available for public review (see *A+ Commission Draft Decision Document*, August 12, 2000).

Summary

The Commission on Student Learning was committed to developing an instructionally relevant, performance-based assessment system that enhances instruction and student learning. The new assessments are based directly on the EALRs. Therefore, teachers and those who provide pre-service and in-service training to teachers should be thoroughly familiar with the EALRs and the assessments that measure them. Teachers and administrators at all grade levels need to be thinking and talking together about what they must do to prepare students to achieve the EALRs and to demonstrate their achievement on classroom-based and state-level assessments.

CRITERION-REFERENCED TESTING

The purpose of an achievement test is to determine how well a student has learned important concepts and skills. Test scores are used to make inferences about students' overall performance in a particular domain. In order to decide "how well" a student has done, some external frame of reference is needed. When we compare a student's performance to a desired performance, this is considered a criterion-referenced interpretation. When we compare a student's performance to the performance of other students, this is considered a norm-referenced interpretation.

Criterion-Referenced Tests are intended to provide a measure of the degree to which students have achieved a desired set of learning targets (desired conceptual understandings and skills) that have been identified as appropriate for a given grade or developmental level in school. Careful attention is given to making certain that the items on the test represent only the desired learning targets and that there are sufficient items for each learning target to make dependable statements about students' degree of achievement related to that target. When a standard is set for a criterion-referenced test, examinee scores are compared to the standard in order to draw inferences about whether students have attained the desired level of achievement. Scores on the test are used to make statements like, "this student meets the minimum mathematics requirements for this class," or "this student knows how to apply computational skills to solve a complex word problem."

Norm-Referenced Tests are intended to provide a general measure of some achievement domain. The primary purpose of norm-referenced tests is to make comparisons between students, schools and districts. Careful attention is given to creating items that vary in difficulty so that even the most gifted students may find that some of the items are challenging and even the student who has difficulty in school may respond correctly to some items. Items are included on the test that measure below-grade-level, on-grade-level, and above-grade-level concepts and skills. Items are spread broadly across the domain. While some norm-referenced tests provide objective-level information, items for each objective may represent concepts skills that are not easily learned by most students until later years in school. Examinee scores on a norm-referenced test are compared to the performances of a norm-group (a representative group of students of similar age and grade). Norm groups may be local (other students in a district or state) or national (representative samples of students from throughout the United States). Scores on norm-referenced tests are used to make statements like, "this student is the best student in the class," or "this student knows mathematical concepts better than 75% of the students in the norm group."

To test all of the desired concepts and skills in a domain, testing time would be inordinately long. Well designed state or national achievement tests, whether norm-or criterion-referenced, always include samples from the domain of desired concepts and skills. Therefore, when state or national achievement tests are used, we generalize from a student's performance on the sample of items in the test and estimate how the student would perform in the domain as a whole. To have a broader measure of student achievement in some domain, it is necessary to use more than one assessment. District and classroom assessments are both useful and necessary to supplement information that is derived from state or national achievement tests.

It is possible, sometimes even desirable, to have both norm-referenced and criterion-referenced information about students' performance. The referencing scheme is best determined by the intended use of the test and this is generally determined by how the test is constructed. If tests are being used to make decisions about the success of instruction, the usefulness of an instructional or administrative program, or the degree to which students have attained a set of desired learning targets, then criterion-referenced tests and interpretations are most useful. If the tests are being used to select students for particular programs or compare students, districts, and states, then norm-referenced tests and interpretations are useful. In some cases, both norm-referenced and criterion-referenced interpretations can be made from the same achievement measures. The *Washington Assessment of Student Learning* (WASL) state level assessment is a criterion-referenced test; therefore, student performance should be interpreted in terms of how well students have achieved the Washington state Essential Academic Learning Requirements.

APPROPRIATE USE OF TEST SCORES

Once tests are administered, WASL performance is reported at the individual, school, and district levels. The information in these reports can be used, along with other assessment information, to help with school and district curriculum planning and classroom instructional decisions. For example, if students in a school are not performing well on the WASL Reading assessment, a careful look at the strand scores (Main Ideas and Details of Fiction; Analysis, Interpretation, & Synthesis of Fiction; Critical Thinking about Fiction; Main Ideas and Details of Non-Fiction; Analysis, Interpretation, and Synthesis of Non-Fiction; Critical Thinking about Nonfiction) can assist in planning instruction in future years. It may be that students as a whole are successful in comprehending and interpreting literature but are not very successful with informational text. Curriculum planning can center on how to improve materials and instruction related to informational text.

While school and district scores may be useful in curriculum and instructional planning, it is important to exercise extreme caution when interpreting individual reports. The items included on WASL tests are samples from a larger domain. Scores from one test given on a single occasion should never be used to make important decisions about students' placement, the type of instruction they receive, or retention in a given grade level in school. It is important to corroborate individual scores on WASL tests with classroom-based and other local evidence of student learning (e.g., scores from district testing programs). When making decisions about individuals, multiple sources of information should be used and multiple individuals who are familiar with the student's progress and achievement (including parents, teachers, school counselors, school psychologists, specialist teachers, and possibly even the students themselves) should be brought together to make such decisions collaboratively.

DESCRIPTION OF THE TESTS

The Grade 10 1999 forms of the Washington Assessment of Student Learning measure students' achievement of the Essential Academic Learning Requirements in Reading, Writing, Listening, and Mathematics. The following tables (Tables 1-1 to 1-4) indicate the EALRs measured by each of the four tests, the test "strands", and the number of items per strand in the 1999 test form.

Table 1-1: 1999 Grade 10, Number and Content of Listening Items

Test Strand*	Number of Items
Listens and observes to gain new information	4
Checks for understanding (paraphrasing, questioning, clarifying)	3
Analyzes media messages	1
Total No. of Items	8

* Listening EALR 1: The student uses listening and observation skills to gain understanding.

Table 1-2: 1999 Grade 10, Number and Content of Reading Items

Type of Reading Passage	Test Strand	Number of Items
Fiction (Literary) ‡	Main ideas, details†	5
	Analyzes, interprets, synthesizes †	8
	Thinks critically*†	6
Non-Fiction (Information or Task Oriented) ‡	Main ideas, details†	8
	Analyzes, interprets, synthesizes †	7
	Thinks critically*†	6
Total Number of Items		40

*Reading EALR 1: The student understands and uses different skills and strategies to read.

†Reading EALR 2: The student understands the meaning of what is read.

‡Reading EALR 3: The student reads different materials for a variety of purposes

Table 1-3: 1999 Grade 10, Number and Content of Writing Prompts

Task	Purposes[1]	Audiences[1]	Process[2]	Number of Prompts	Scores[3]
Extended Piece	Persuade	Editor	• prewrite • first draft • revise • edit • final draft	1	• Content, Organization & Style • Writing Mechanics
Extended Piece	Inform	Fellow Student	• prewrite • first draft • revise • edit • final draft	1	• Content, Organization & Style • Writing Mechanics
Total Number of Prompts				2	

[1] Writing EALR 1: The student writes clearly and effectively (concept & design, style [word choice, sentence fluency, voice], and conventions).

[2] Writing EALR 2: The student writes in a variety of forms for different audiences and purposes.

[3] Writing EALR 3: The student understands and uses the steps of a writing process*

Table 1-4: 1999 Grade 10, Number and Content of Mathematics Items

Process Strand	Concept Strand	Number of Items
Concepts & Procedures	Number Sense[1]	7
	Measurement[1]	6
	Geometric Sense[1]	6
	Probability and Statistics[1]	6
	Algebraic Sense[1]	6
Solves Problems[2]		3
Reasons Logically[3]		4
Communicates Understanding[4]		4
Making Connections[5]		4
Total No. of Items		46

[1] Mathematics EALR 1: The student understands and applies the concepts and procedures of mathematics.

[2] Mathematics EALR 2: The student solves problems using mathematics.

[3] Mathematics EALR 3: The student uses mathematical reasoning.

[4] Mathematics EALR 4: The student communicates knowledge and understanding in mathematical and everyday language.

[5] Mathematics EALR 5: The student makes mathematical connections.

ESTIMATED TESTING TIME PER SESSION—10th GRADE - SPRING 1999

The tests in the *Washington Assessment of Student Learning* are not timed. Students should have as much time as they need to work on the tests. Professional judgment should determine when a student is no longer productively engaged. When the majority of students have finished, the few still working may be moved to a new location to finish. Teachers' knowledge of students' work habits or special needs may suggest that some students who work very slowly should be tested separately or grouped with similar students for the entire assessment. For planning purposes, the estimated testing times required for most students are given in Table 1-5.

Table 1-5: Estimated Testing Times for Grade 10 WASL

Session	Subject	Approximate Time[1]
1	Listening	25 minutes
	Reading (Day One)	60 minutes
2	Reading (Day Two)	40 minutes
	Writing (Day One)	75 minutes
3	Writing (Day Two)	75 minutes
4	Mathematics (Day One) with tools	80 minutes
5	Mathematics (Day Two) without tools	80 minutes

[1] Above times are estimates for actual testing time. Additional time will be required to distribute and collect materials and cover the directions for test-taking. Testing sessions need not follow on consecutive days. Individual sessions should not be split but may be spaced with one or more days in between.

PART 2

TEST DEVELOPMENT AND CONTENT REPRESENTATION

The content of the *Washington Assessment of Student Learning* (WASL) state assessment is derived from the Washington state Essential Academic Learning Requirements (See Appendix A for an overview). These Essential Academic Learning Requirements (EALRs) define, for Washington schools, what students should know and be able to do by the end of grades 4, 7, and 10 in Reading, Writing, Communication, Mathematics, and by the end of grades 5, 8, and 10 in history, geography, economics, civics, science, the arts, health, and fitness. The 1999 WASL tests measured EALRs for Reading, Writing, Mathematics, and Listening in grades 4, 7 and 10.

ITEM AND TEST SPECIFICATIONS

The first step in the test development process was to select the "Content Committees" that worked with staff of the Commission on Student Learning (CSL) and the Contractor (Riverside Publishing Company) to develop the actual items, which make up the assessments at each grade level. Each Content Committee was composed of 20 to 25 persons from around the state, most of whom were classroom teachers and curriculum specialists who had teaching experience at or near the grades and in the content areas that were to be assessed (i.e., Reading/Writing/Communication or Mathematics).

The second step in the development process was coming to a common agreement about the meaning and interpretation of the EALRs as well as which ones could be assessed on the state level test. Here it was very important that the Contractor, the Content Committees and the CSL staff were in agreement, in concrete ways, about what students were expected to know and be able to do and how these skills and knowledge would be assessed. In addition, the benchmark indicators were combined in various ways to create testing **targets** for which items would be written (See Appendix B and C).

Next, test specifications were prepared. Test specifications define and describe such details as the kinds and number of items on the assessment, the blueprint or physical layout of the assessment, the amount of time to be devoted to each content area, and the scores to be generated once the test is administered. It was important that the goals of the assessment and the ways in which the results would be used be established at this stage so that the structure of the test would support the intended uses. In addition, the Test Specifications are the basics for developing equivalent test forms in subsequent years as well as creating new items to supplement the item pool. The final Test specifications (See Appendix B and C) document the following topics:

- Purpose of the Assessment
- Strands
- Item Types
- General Considerations of Testing Time and Style
- Test Scoring
- Distribution of Test Items by Item Type

There are three types of items on the *Washington Assessment of Student Learning* (WASL) tests: multiple choice, short answer, and extended response. For each multiple-choice item, students select the one best answer from among three or four choices provided. Each multiple- choice item is worth one point. These items are machine scored.

The other two "open-ended" item types—short answer and extended response—require students to give their own response in words, numbers, or pictures (including graphs or charts). Short-answer items are worth two points (scored 0, 1, or 2) and extended-response items are worth four points (scored 0, 1, 2, 3, or 4). For these items, student responses are assigned partial or full credit based on carefully defined scoring criteria. These items cannot be scored by machine and require hand-scoring by well-trained professional scorers (See Part 4).

In addition to the three item types, students are asked to do two writing assignments (prompts). For grade 10, students write one informative piece and one persuasive piece. The writing prompts may require students to write a letter requesting information, describe an important event or situation, explain a procedure for completing a task or project. etc. Each written piece is worth six points and is hand-scored for content, organization, and style (1, 2, 3, or 4 points) and mechanics and spelling (0, 1, or 2 points).

Tables 2-1 through 2-3 are the test blueprints for item content and item types for the Reading, Listening, and Mathematics tests of the Grade 10 test.

Table 2-1: Grade 10 Reading Test: Item distribution by text type, strand, and item type

Text types/Strands	No. of Reading Selections	No. of Words Per Passage	No. of Multiple-Choice Items	No. of Short Answer Items	No. of Extended Response Items
Fiction‡	3	up to 1300	10-15	3-6	1
Comprehends important ideas and details†			3-5	1-2	0
Analyzes, interprets, synthesizes†			2-5	1-2	0-1
Thinks critically†*			2-5	1-3	0-1
Non-fiction‡	3-4	up to 1300	10-15	3-6	1
Comprehends important ideas and details†			3-5	1-2	0
Analyzes, interprets, synthesizes†			2-5	1-3	0-1
Thinks critically†*			2-5	1-3	0-1
Total	6-7	up to 4000	26-30	9-11	2

*Reading EALR 1: The student understands and uses different skills and strategies to read.

†Reading EALR 2: The student understands the meaning of what is read.

‡Reading EALR 3: The student reads different materials for a variety of purposes

Table 2-2: Grade 10 Listening Test: Item distribution by strand and item type

Strands	Number of Reading Selections	Number of Words Per Passage	Number of Multiple-Choice Items	Number of Short Answer Items
	2 editorials	up to 100	6-8	2
Listens and observes to gain and interpret information			3-5	0
Checks for understanding			2-3	1
Analyzes media messages			0-1	1
Total	2 editorials	up to 200	6-8	2

* Listening EALR 1: The student uses listening and observation skills to gain understanding.

Table 2-3: Grade 10 Mathematics Test: Item distribution by strand and item type

Strands	Multiple Choice	Short Answer	Extended Response
Number Sense[1]	3-7	1-2	0
Measurement Concepts[1]	3-7	1-2	0
Geometric Sense[1]	3-7	1-2	0
Probability and Statistics Procedures[1]	3-7	1-2	0
Algebraic Sense[1]	3-7	1-2	0
Solves Problems[2]	0-2	2-4	1-2
Reasons Logically[3]	0-2	1-4	0-1
Communicates Understanding[4]	0-2	1-4	0-1
Making Connections[5]	0-2	1-4	0-1
Maximum Number of Items	30	12	4
Maximum Number of Points	30	24	16

[1]Mathematics EALR 1: The student understands and applies the concepts and procedures of mathematics.

[2]Mathematics EALR 2: The student solves problems using mathematics.

[3]Mathematics EALR 3: The student uses mathematical reasoning.

[4]Mathematics EALR 4: The student communicates knowledge and understanding in mathematical and everyday language.

[5]Mathematics EALR 5: The student makes mathematical connections.

Based on the clarification of the EALRs and the Test Specifications, the next step was to develop Item Specifications. Item specifications provide sufficient detail, including sample items, to direct item writers in the development of appropriate test items for each assessment strand. Separate specifications were produced for the different item types including multiple-choice, short answer and extended response. The Test and Item Specification documents were not only essential for WASL test construction but taken together they are powerful tools for teachers in developing their own assessments and for administrators in reviewing instructional programs. Test and Item Specifications are updated yearly, as needed. The most recent versions of these specifications can be obtained through the web site (www.k12.wa.us) for the Washington State Office of the Superintendent of Public Instruction (OSPI).

CONTENT REVIEWS

Once the Test and Item Specifications were completed and reviewed by the Content Committees, the Contractor's item writers prepared sample items and scoring criteria based on these specifications. The Content Committees task was then to review the items and scoring criteria to assure that the item writers had followed the specifications. As necessary items were revised to ensure that they measured Washington's Essential Academic Learning Requirements both accurately and comprehensively.

When the Content Committees were satisfied that the sample items and scoring criteria were appropriate, the item writers then produced literally hundreds of items to be pilot tested at the selected grade levels. Each test item was coded by content (EALR) area and item type (multiple choice, short answer, extended response) and presented to the Content Committees for final review just as they were to appear on the pilot test forms (including graphics, art work, and location on pages).

When the draft items were completed, the Content Committees reviewed each item, focusing on its fit to the Item Specifications, the EALRs, and the appropriateness of item content. For all short answer and extended response items, the proposed scoring guidelines (rubrics) were also reviewed. The Committees had three options with each item: approve the item (and scoring guidelines) as presented, recommend changes or actually edit the item (or scoring guidelines) to improve the item’s "fit" to the EALRs and the Specifications, or eliminate the item from use in the assessment.

In addition to the Content Committees, a separate Fairness Review Committee reviewed each item to identify language or content that might be inappropriate or offensive to students, parents, or communities or items which might contain "stereotypic" or biased references to gender, ethnicity, or culture. As with the Content reviews, The Fairness Review Committee reviewed each item and accepted, edited, or rejected it for use on the pilot assessment.

In order to be included on the pilot assessment, every item was reviewed by both the Content Committees and the Fairness Review Committee. Approved items were to:

- be appropriate measures of the intended content;
- be appropriate in difficulty for the grade level of the examinees;
- have only one correct or best answer for each multiple-choice item;
- have appropriate and complete scoring guidelines for the open response items
- be free from content that might disadvantage some students for reasons unrelated to the concept or skill being tested

ITEM TRYOUTS

The approved items were then assembled into pilot test forms and administered to carefully-selected, representative samples of students across the state. All schools in the state of Washington were invited to participate in the pilot testing. Eighty five percent of fourth graders took part in the pilots. Test forms were randomly distributed with some effort to ensure that each test form was administered in districts with high populations of ethnic minority students. Each test form was administered to at least 1000 students.

SCORING AND ITEM ANALYSIS

Following the administration of the pilot assessment, student responses were evaluated by applying the scoring criteria approved by the Content Committees. A variety of statistical analyses were then employed to determine the effectiveness of the items and to check for item bias that may have been missed by the earlier reviews.

Two methods were used for item analysis. These were traditional or classical item analysis, which included the item means and item-test correlations for each item, and Rasch analysis, which included the item location and item fit. In addition, bias analysis was conducted using the Mantel-Haenszel bias statistic. Bias analysis investigates whether there is differential item performance for examinees of the same abilities who differ by virtue of gender or ethnicity.

Rasch Analysis

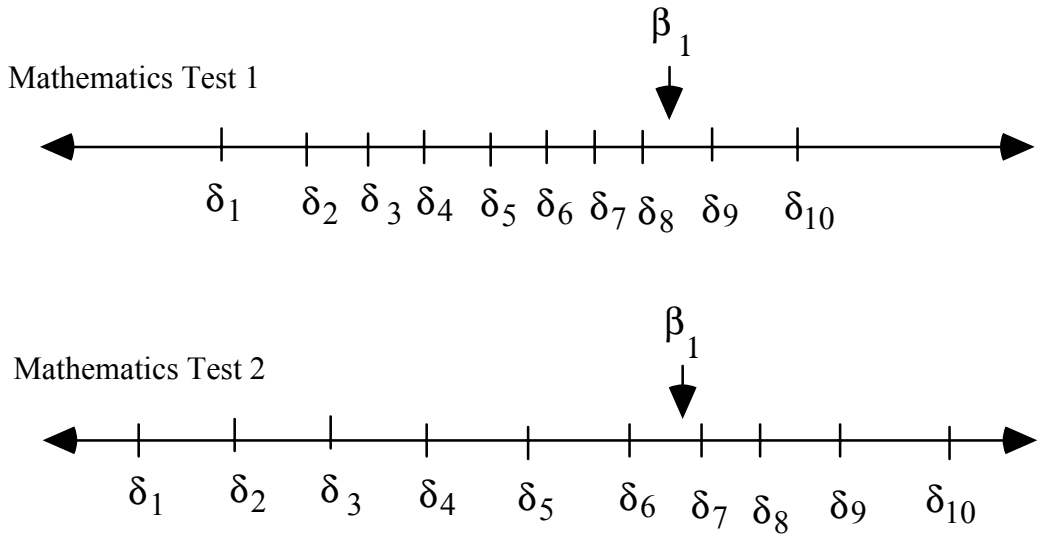
Rasch analysis is an Item Response Theory (IRT) analysis that places all items on a unique continuous scale for each content area. In addition, all examinees in the tryout pool are located on the same underlying scale. The Rasch analysis process separates item difficulty parameters from the abilities of the examinees in the sample that was tested. In this way, item difficulty parameters can be assumed to be the same for groups who are different from the original sample. The basic formula for the Rasch model is:

$$p[x_{vi} = 1 \mid \beta_v, \delta_i] = \frac{\exp(\beta_v - \delta_i)}{1 + \exp(\beta_v - \delta_i)}$$

Where p = the probability of getting an item right given the ability of the examinee (βv) and the difficulty of the item (δi).

Working from this formula, item difficulties and examinee abilities can be estimated for a given test. The item difficulty location is the point on the ability scale where examinees have a 50/50 chance of getting an item correct. Figure 2-1 shows how examinee ability and item difficulty are placed on ability scales.

Figure 2-1: Location of examinee β_1 on two tests with item difficulties δ_1 through δ_{10}



Because the Rasch model can obtain an equal interval scale independent of item difficulty and person performance, the meaning of test scores can be interpreted in terms of scaled scores rather than number correct scores. For example, in Figure 2-1 (above), the examinee (β_1) got the first eight items correct on Mathematics Test 1 and the first six items right on Mathematics Test 2. The examinee is the same and her/his mathematics knowledge and skill remains the same; however, the ease or difficulty of the items result in different number-correct scores. The Rasch model will indicate the true distance of items from one another across the scale so that examinee test scores reflect the relative distance along the scale rather than the number of items answered correctly. The Rasch model separates item difficulty from examinee ability so that scores of examinees can be interpreted in terms of an underlying ability scale.

For items that have multiple points, a partial credit Rasch model is used to estimate the difficulty (threshold) of each *score* for an item. For example, items with 2 possible points can have two item thresholds: one for the point on the scale (location) at which examinees with abilities equal to that level on the scale have an equal chance of getting a score of 0 or 1, and one for the point on the scale at which examinees with abilities equal to that level on the scale have an equal chance of getting a score of 1or 2. The formula for Master's partial credit model (which uses the Rasch dichotomous model as its base) is:

$$\pi_{xvi} = \frac{\exp \sum_{j=0}^{x} (\beta_v - \delta_{ij})}{\sum_{k=0}^{m_i} \exp \sum_{j=0}^{k} (\beta_v - \delta_{ij})} \qquad x = 0, 1, \ldots m_i$$

Where π equals the probability that an examinee with ability β_n will get score *x* on item *i* and δ_{ij} is the location of "step" *j* for item *i* (the point on the underlying scale where examinees have an equal probability of getting two adjacent scores [e.g., a score of 0 and a score of 1] on the item).

Once item scores are placed on a scale, items are assessed for "fit" to the Rasch model. The Rasch model assumes there was no guessing on multiple choice items and that, even though the items differ in terms of difficulty (or location on the scale) the items all function equally in discriminating between examinees below and above a given location on the scale. In order to be retained in the item pool, items must measure relevant knowledge and skill, represent desired locations on the ability scale, and fit the Rasch model.

Rasch analyses were conducted independently for each content area within the Washington Assessment of Student Learning (WASL). The fit of items depends upon whether the items in a scale were all measuring a similar body of knowledge and skill—in other words, whether the scale was unidimensional. Just as height, weight, and body temperature are different dimensions of the human body, so are Reading, Writing, Mathematics, and Listening different dimensions of achievement. Therefore, the items and scales for each test are examined independently.

In order to place all items across test forms on the same Rasch scale, a subset of items was repeated in adjacent forms. In other words, five items in Form 1 were repeated in Form 2; a different five items in Form 2 were repeated in Form 3; a different five items in Form 3 were repeated in Form 4; a different five items in Form 4 were repeated in Form 5; a different five items in Form 5 were repeated in Form 6; a different five items in Form 6 were repeated in Form 7; a different five items in Form 8 were repeated in Form 1. In this way, Form 1 could be the anchor form and all items could be calibrated back to the item locations for the items in Form 1.

Traditional Item Analysis

For multiple-choice items, item means and item-test correlations constitute p-values and point-biserials respectively. These are the classical test theory equivalent of item difficulties and item discriminations. The p-value tells the percent of examinees who responded correctly to an item. Its value can range from 0 to 1.0. The point-biserial gives a measure of the relationship between performance on an item and performance on the test as a whole and can range from -1.0 to 1.0. For multiple-point (open-ended items), item means indicate the average earned score for examinees in the tryout sample. For 2-point items, item means can range from 0 to 2. For four-point items, item means can range from 0 to 4. Item-test correlations, for multiple point items, indicate the relationship between item performance and test performance. Item-test correlations can range from -1.0 to 1.0. Item-test correlations are computed using the test scores relevant to the item.

Unlike the Rasch item data, item means and item-test correlations are dependent on the sample of examinees that took the various tests. If the examinees were exceptionally well schooled in the concepts and skills tested, item means will be fairly high and the items will appear to be easy.

If examinees are not well schooled in the concepts and skills tested, item means will be fairly low and items will appear to be difficult. If performance on an item does not relate well to performance on the test as a whole, item test correlations will be low or even negative. Hence both Rasch data and traditional item analysis data are used in item selection.

Bias Analysis

The Mantel Haenszel statistic is a chi-square (χ^2) statistic. Examinees are separated into relevant groups based on ethnicity or gender. Examinees in each group are ranked in terms of their total score on the relevant test. Examinees in the focal group (e.g., females) are compared with examinees in the reference group (e.g., males) in terms of their performance on individual items. Multiple 2x2 tables are created for each item (one for each total test score) indicating, for that score, the number of examinees in each group who got the item right and the number of examinees in each group who got the item wrong. Table 2-4 shows an example 2x2 table for performance on a hypothetical item for males and females with a total test score of 10 on a 40 point test. It appears that the item is more difficult for females than it is for males who had a total test score of 10.

Table 2-4: Responses to Item 3 for Males and Females with Total Test Score of 10

Item Number 3	Number Responding Correctly	Number Responding Incorrectly
Males (N = 100)	50	50
Females (N = 100)	30	70

Examinees with Total Test Score = 10

To complete the Mantel-Haenszel statistic, similar 2x2 tables are created for every test score. A χ^2 statistic is computed for each 2x2 table and the sum of all of the χ^2 statistics across all test scores gives the total bias statistic for a single item. When items have multiple points, a generalized Mantel-Haenszel statistic is computed using all points. Items that demonstrate a high $\Sigma\chi^2$ are flagged for potential bias. Generally, a certain percent of the items in any given pool of items will be flagged for item bias by chance alone. Careful review of items can help to identify whether some characteristic of an item may cause the bias (e.g., the content or language is unfamiliar to girls) or whether the bias data is probably a result of statistical error. For the WASL analyses, the alpha level (error level) was set at .01; therefore, about 1 percent of the items are expected to be flagged for bias by chance alone.

ITEM SELECTION

Statistical review of items involves examining item means, Rasch item difficulties (locations on the ability scale), and item-test correlations to determine whether items are functioning well. In addition, statistical review requires examining the "fit" of items to the Rasch model. Items that have extremely poor fit to the Rasch model must be revised or removed from the item pool prior to building a final test form. Items that function very poorly (are too easy, too difficult, or have low or negative item-test correlations) must also be revised or removed from the item pool. Finally, items that are flagged for bias against a focal group are examined closely to decide whether they will be removed from the pool. Generally, when item tryouts are conducted, sufficient numbers of items are developed so that revision and new tryouts are not needed. Faulty items can be deleted from the item pool.

After the statistical analyses were completed for the WASL, the Content and Fairness Review Committees reviewed these results and made the final determination about item quality and appropriateness based on the pilot test data. Items were reviewed again for fit to the EALRs; scoring rules were reviewed again for fit to the EALRs and to the demands of the items. In the Fairness Review Committees, bias data were reviewed to determine whether content or language may have resulted in large bias statistics. During these reviews, items were either accepted or rejected for the final pool of items.

Once these reviews were completed, the final pool of items was used to develop "operational" test forms. Operational test forms are those that are administered each year to monitor progress of schools and districts in helping students achieve the EALRs. Each operational form is developed by selecting items from the large pool of items tested in the 1998 item tryouts and approved by the Content and Fairness Review Committees. Four criteria are used in item selection for test forms:

1 Item quality
2 Content representation (See Test Specifications)
3 Representation of all gender and ethnic groups (See Test Specifications)
4 Item locations

Item quality is determined by the item means, item-test correlations, bias statistics, Rasch item locations, and fit statistics. Only the best items from the final pool are to be used in the operational test forms. Test specifications guide item selection to ensure that all relevant strands are represented in each test form as defined in the Test Specifications. Representation of all gender and ethnic groups is reviewed to ensure that Reading and Listening passages and stimulus materials used in the Mathematics and Writing tests give balanced representations of groups. Finally, because the WASL is intended to be a criterion-referenced test, and because performance standards are established for each test, item have been selected to represent a range of locations on the Reading, Mathematics, Writing, and Listening scales. After proficiency scores were established for each test in 1999 (See Part 5), item selection for subsequent years has ensured that item locations are similar to those in the initial operational test form in 1999.

Following the administration of the first operational Grade 10 assessment in Spring of 1999, the tests were scored for all participating students. A Standard-Setting Committee (see Part 5) was convened to establish the performance levels appropriate for reporting students' achievement of the EALRs. Based on the standards set by the Committee and approved by the Commission on Student Learning, results for the first Grade 10 operational assessment were reported in September, 1999. Table 2-5 gives the schedule of test development that was used for the Grade 10 WASL.

Table 2-5: Test Development Process for Grade 10

Action	Dates
Essential Academic Learning Requirements	March 1995
Test and Item Specifications	July-August 1997
Item Development	Sept.- Oct. 1997
Item Review (Content and Fairness)	November 1997
Pilot Testing	May 1998
Item Review (Content and Fairness)	Aug 1998
Item Bank	Sept 1998
Score Reports Designed	Sept 1998
Operational Tests Created	Oct - Dec 1998
Published Example Test Assessment Sampler	Feb 1999
First Operational Test Administered	April - May 1999
Standard Setting	June 1999

PART 3
EVIDENCE FOR THE VALIDITY OF INFERENCES FROM TEST SCORES

The most important issue in test development is the degree to which the achievement test actually elicits the conceptual understanding and skills that it is supposed to measure. In other words, when one claims that students must use logical reasoning skills to respond to an item, we need evidence that logical reasoning rather than memorization was actually used in the students' responses. Validity is an evaluative judgment about the degree to which the test ***scores*** can be interpreted to mean what test developers claim that they mean. Generally, there are about a half dozen different strategies for obtaining evidence for the validity of test scores (Messick, 1989):

1. We can look at the content of the test in relation to the content of the domain of reference;
2. we can probe the ways in which individuals respond to the items or tasks;
3. we can examine the relationships among responses to the tasks, items, or parts of the test, that is, the internal structure of test responses;
4. we can survey relationships of test scores with other measures and background variables, that is, the test's external structure;
5. we can investigate differences in these test processes and structures over time, across groups and settings, and in response to . . . interventions such as instructional . . . treatment and manipulation of content, task requirements, or motivational conditions;
6. finally, we can trace the social consequences of interpreting and using test scores in particular ways, scrutinizing not only the intended outcomes, but also the unintended side effects. (p. 16)

Validity, then, is a multidimensional construct that resides, not in tests, but in the relationships between any test score and its context (including the instructional practices and the examinee), the knowledge and skills it is to represent, the intended interpretations and uses, and the consequences of its interpretation and use. Messick stated that multiple sources of evidence are needed to investigate the validity of assessments. The following pages provide a description of the evidence available for the validity of scores on the Grade 10 *Washington Assessment of Student Learning* (WASL). This includes: correlations among scores and strands within the WASL and factor analysis studies examining evidence for the construct validity of WASL.

Part 2 of this technical report describes the process used in relation point 1 above: the judgment of content in relation to the subject area domains and selection of items that have adequate psychometric characteristics. While content representation and item quality are important aspects of tests, they do not ensure the validity of test scores. In order to examine the validity of test scores, it is important to determine whether examinees' performance within the set of items on the test is consistent (internal structure). This type of evidence is considered evidence for the construct validity of test scores. Studies to examine internal structure question whether the test scores elicit the constructs (knowledge and skills) the tests were intended to elicit.

Studies have been conducted to gather evidence for the construct validity of the WASL Grade 10 Reading, Writing, Listening, and Mathematics Tests. The internal structure of the tests was examined by looking at the intercorrelations among the tests and strands assessed by the test. It will be useful to compare the performance presented in Part 8 of each year's Grade 10 WASL technical reports to determine whether the scores for these tests are increasing over time. Since the test forms are different each year, improvement in scores over time would suggest that performance on the

WASL tests are affected by instruction aimed at the Washington state EALRs. The first year for the administration of the Grade 10 WASL was 1999; therefore, state trend data are not yet available.

INTERNAL EVIDENCE FOR THE VALIDITY OF WASL SCORES

Correlations Among WASL Test Scores

The first analysis was that of correlations among WASL test scores. As can be seen in Table 3-1, responses to the different tests are moderately to strongly related. The strongest correlation is between scores on the WASL Reading Test and scores on the WASL Mathematics Test (.727). The next strongest are correlations between WASL Reading scores and WASL Writing scores (.646) and WASL Reading scores and WASL Listening scores (.649). Performance on the Listening Test was moderately correlated with performance on the Writing and Mathematics Tests. WASL Mathematics scores were moderately related to WASL writing scores.

Table 3-1: 1999 Grade 10 Correlations among WASL Test Scores

Tests	WASL Listening	WASL Reading	WASL Writing	WASL Mathematics
WASL Listening	1.00	.649	.512	.559
WASL Reading		1.00	.646	.727
WASL Writing			1.00	.589
WASL Mathematics				1.00

Intercorrelations among WASL Strand Scores

To more closely examine the relationships among performances on the WASL tests, the second analysis was of correlations among strand scores for Reading, Mathematics, and Writing, as well as the Listening Test scores. Table 3-2 gives the correlations among the strands within the 1999 WASL. As can be seen, scores for Reading strands (Main Ideas and Details of Fiction, Analysis, Interpretation, and Synthesis of Fiction, and Critical Thinking about Fiction, Main Ideas and Details in Nonfiction Text, and Analysis, Interpretation, and Synthesis of Nonfiction Text, Critical Thinking about Nonfiction Text) are moderately well correlated (.540 to .679) with the strongest correlations between strands that measure analysis, interpretation, and synthesis of text and thinking critically about for all types of text. The Writing Content, Organization, & Style score is well correlated with the Writing Mechanics score (.607). Correlations among the Mathematics concepts scores (Number Sense, Measurement, Geometric Sense, Probability and Statistics, and Algebraic Sense) are moderately well correlated as would be expected given that these are diverse conceptual areas of Mathematics (.563 to .630). Prior research has shown that students perform differently on mathematical tasks that tap different areas of mathematics (Shavelson, Baxter, & Gao, 1993). The highest correlation is between Number Sense and Algebraic Sense (.630). Given that facility with numbers is required for both strands, this is to be expected.

Correlations among the Mathematical process scores (Solves Problems, Reasons Logically, Communicates Understanding, and Makes Connections) are also moderately strong (.633 to .682). The highest correlation is between scores for Solves Problems and scores for Makes Connections (.682). Since items for the Solves Problems strand are situated in contexts and require multiple content strands, this fairly high correlation is expected. The next highest correlation is between

Solves Problems and Reasons Logically. It is likely that students must use many of their logical reasoning skills to solve problems.

Correlations between Mathematics content scores and Mathematics process scores are informative. Scores for Solves Problems, Reasons Logically, Communicates Understanding, and Makes Connections are moderately well correlated with scores for all content strands scores (.588 to .687). This suggests that content understandings are required for successful performance on all of the process strands. The highest correlations are between Solves Problems and Algebraic Sense (.687) and between Makes Connections and Algebraic Sense (.666). Given the strong relationship between Makes Connections and Solves Problems as well as the age of the students, these relationships make sense.

Correlations between Reading strand scores and Mathematics content strand scores are low to moderate (.415 to .551) with most between .40 and .50. The correlations between Reading strand scores and Mathematics process strand scores are also low to moderate (.464 to .580). The strongest relationships are between Thinks Critically about Fiction and all Mathematics strand scores (.497 to .580). It is important to note that correlations between Writing strand scores and Mathematics strand scores are moderately low (.368 to .477). Writing strand scores also have only moderate correlations with all Reading strands with most between .40 and .50. These correlations suggest that, for both the Reading Test and the Mathematics Test, skill in writing is only moderately related to performance.

Table 3-2: 1999 Grade 10 Correlations among Strands in the WASL

Strands	RL2	RL3	RI1	RI2	RI3	W1	W2	NS	ME	GS	PS	AS	SP	RL	CU	MC
Ideas & Details Fiction	.584	.658	.540	.573	.565	.504	.453	.453	.443	.435	.497	.466	.497	.475	.502	.474
Interpretation Fiction		.639	.553	.572	.556	.477	.436	.432	.430	.415	.485	.438	.467	.458	.486	.449
Thinks Critically about Fiction			.575	.679	.640	.582	.496	.510	.497	.500	.551	.537	.580	.545	.575	.552
Ideas & Details Nonfiction				.611	.609	.423	.395	.432	.447	.418	.484	.448	.481	.464	.488	.453
Interpretation Nonfiction					.654	.522	.457	.467	.466	.442	.514	.483	.518	.488	.527	.507
Thinks Critically about Nonfiction						.476	.425	.452	.461	.429	.497	.462	.504	.481	.515	.479
Content, Organization, Style							.607	.439	.412	.414	.466	.453	.477	.446	.477	.471
Writing Mechanics								.394	.368	.373	.422	.394	.410	.381	.419	.407
Number Sense									.617	.569	.624	.630	.634	.597	.605	.626
Measurement										.563	.619	.616	.640	.601	.618	.612
Geometric Sense											.585	.601	.617	.588	.594	.588
Prob. & Statistics												.623	.642	.622	.654	.630
Algebraic Sense													.687	.638	.617	.666
Solves Problems														.659	.647	.682
Reasons Logically															.640	.633
Communicates																.637

RL1-Main Ideas & Details of Fiction
RL2-Analyzes and Interprets Fiction
RL3-Thinks Critically about Fiction
RI1-Main Ideas & Details of Nonfiction
RI2-Analyzes and Interprets Nonfiction
RI3-Thinks Critically about Nonfiction
W1-Content, Organization, & Style
W2-Writing Mechanics

NS-Number Sense
ME-Measurement
GS-Geometric Sense
PS-Probability and Statistics
AS-Algebraic Sense

SP-Solves Problems
RL-Reasons Logically
CU-Communicates Understanding
MC-Makes Connections

Factor Analysis of WASL Listening Test Scores and Reading, Writing, and Mathematics Strand Scores

In order to follow up on these correlations, an exploratory factor analysis was conducted with the Listening Test scores, and the Writing, Mathematics and Reading strand scores. A principal components analysis was conducted using SPSS 8.0. The number of factors was determined using three criteria: eigenvalues greater than one, a scree test, and the solution in which at least 63 percent of the variance was explained. Varimax (orthogonal) rotation was used. There were two plausible factor structures in the data. One (using the eigenvalues criterion) resulted in a two-factor solution that explained 55 percent of the total variance. Using a criterion of .60 for factor loadings (36% of the variance of a given variable), the two underlying factors were the language arts (Listening, Reading and Writing) and Mathematics. Table 3-3 gives the factor loadings (correlations between each of the variables and the underlying factors) from the rotated component matrix for the two-factor solution.

Table 3-3: 1999 Grade 10 Rotated Factor Loadings for Listening, Reading, Writing and Mathematics Strands for Two-Factor Solution

Variables	Language Arts Factor	Mathematics Factor
Listening	.694	.289
Main Ideas and Details of Fiction	.724	.309
Analysis and Interpretation of Fiction	.733	.273
Critical Thinking about Fiction	.755	.391
Main Ideas and Details of Nonfiction	.691	.305
Analysis and Interpretation of Nonfiction	.746	.326
Critical Thinking about Nonfiction	.732	.311
Content, Organization, and Style in Writing	.644	.324
Writing Mechanics	.611	.261
Number Sense	.301	.748
Measurement	.290	.751
Geometric Sense	.282	.726
Probability and Statistics	.390	.715
Algebraic Sense	.301	.777
Solves Problems	.354	.771
Reasons Logically	.338	.740
Communicates Understanding	.405	.710
Makes Connections	.333	.760

The second analysis, using percent of variance explained and the scree plot as the criteria, resulted in a three-factor solution that explained 68 percent of the total variance. Using a criterion of .60 for factor loadings (36% of the variance of a given variable), the underlying factors were Reading and Listening, Mathematics, and Writing. Table 3-4 gives the factor loadings (correlations between each of the variables and the underlying factors) from the rotated component matrix for the three-factor solution. While Reading, Writing, and Mathematics may be moderately correlated,

Listening/Reading, Writing, and Mathematics strands represent separate dimensions of performance on the 1999 WASL as a whole. The fact that the Listening test scores load on the same factor as all Reading strand scores (albeit the lowest loading for that factor - .642) probably reflects the general comprehension required for both tests.

Table 3-4: 1999 Grade 10 Rotated Factor Loadings for Listening, Reading, Writing and Mathematics Strands for Three-Factor Solution

Variables	Mathematics Factor	Listening and Reading Factor	Writing Factor
Listening	.282	.642	.274
Main Ideas and Details of Fiction	.302	.677	.272
Analysis and Interpretation of Fiction	.269	.713	.219
Critical Thinking about Fiction	.383	.693	.312
Main Ideas and Details of Nonfiction	.307	.750	.049
Analysis and Interpretation of Nonfiction	.321	.730	.216
Critical Thinking about Nonfiction	.309	.754	.134
Content, Organization, Style in Writing	.298	.356	.745
Writing Mechanics	.232	.278	.826
Number Sense	.743	.253	.188
Measurement	.748	.280	.108
Geometric Sense	.721	.242	.166
Probability and Statistics	.710	.351	.190
Algebraic Sense	.772	.259	.178
Solves Problems	.766	.318	.177
Reasons Logically	.737	.319	.138
Communicates Understanding	.705	.370	.185
Makes Connections	.754	.283	.200

PERFORMANCE ACROSS GROUPS

Part 8 of this technical report presents data regarding performance of examinees across different categorical programs (i.e., Title I Reading, Title I Mathematics, LAP Reading, LAP Mathematics, S504, Special Education, Highly Capable Students, Bilingual/ESL, Title I Migrant). These data can be examined to determine whether patterns of performance, are what would be expected on the test based on examinees' special needs. For example, students who have been identified as "highly capable" outperform all other groups on all tests. In addition, students who are in Title I Migrant and Bilingual/ESL programs have difficulty with reading and writing performance. Gender groups are also compared in Part 8. Whereas boys and girls perform equally well in Mathematics and Reading, girls outperform boys in Listening and Writing. These data, and other patterns in Tables 8-3 through 8-14, suggest that scores on the WASL tests are consistent with other measures of achievement in these subject areas.

SUMMARY

The results of these analyses provide evidence to support the validity of 1999 WASL scores. While achievement in one subject area is generally related to achievement in other subject areas, once WASL subscores are examined, it is evident that Listening and Reading, Mathematics, and Writing are different underlying dimensions of performance on the WASL tests.

Reference

Shavelson, R. J., Baxter, G. P., Gao, X. (1993). Sampling variability of performance assessments. *Journal of Educational Measurement, 30*, 215-232.

PART 4

SCORING THE WASL OPEN-ENDED ITEMS

During item development, scoring criteria for each open-ended item on the *Washington Assessment of Student Learning* (WASL) were written. Appendix D provides the general scoring criteria that served as the guides for the item specific scoring criteria for Reading, Mathematics, and Listening items. Appendix D also provides an example of how the general scoring criteria for a mathematics item was made specific to the requirements of the task. During item reviews, the scoring criteria were reviewed along with item directions. A central aspect of the validity and reliability of test scores is the degree to which scoring criteria are related to the appropriate learning targets (Essential Academic Learning Requirements) and whether they are applied faithfully during scoring sessions. Appendix D also provides the scoring criteria for all student writing samples. The following procedures were used to score the WASL items and apply to all content areas that include open-ended questions calling for student-constructed responses. These procedures were used for the full pool of items that were pilot tested as well as for the 1999 operational tests.

QUALIFICATIONS OF READERS

Highly-qualified, experienced readers (scorers) were essential to achieving and maintaining consistency and reliability when scoring student-constructed (open-ended) responses. Readers selected for the Washington Assessment of Student Learning were required to have the following qualifications:

- A minimum of a bachelor's degree in an appropriate academic discipline (such as English, English Education, Math, Math Education, or related fields);
- Demonstrable ability in performance assessment scoring;
- Teaching experience, especially at the elementary or secondary level, was preferred.

Team and table leaders, responsible for supervising small groups of readers, were selected on the basis of demonstrated expertise in all facets of the scoring process, including strong organizational abilities, leadership, and interpersonal communication skills.

RANGE-FINDING AND ANCHOR PAPERS

The thoughtful selection of papers for range-finding and the subsequent compilation of anchor papers and other training materials were the essential first steps to ensure that scoring was conducted consistently, reliably, and equitably.

In the range-finding process, performance assessment and curriculum specialists working with team and table leaders and teachers from Washington all became thoroughly familiar with and reached consensus on the scoring criteria (rubrics) approved by the Content Committees for each open-ended item. These range-finding teams began work with random selections of student responses for each item. They reviewed these responses, selected an appropriate range of responses, and placed them into packets, numbered for easy reference. The packets of responses were read independently by members of a team of the most experienced readers. Following these independent readings and tentative ratings of the papers, the total range finding group worked together to discuss both the common and divergent scores. From this work, they assembled tentative sets of example responses for each prompt.

The primary task of the range-finding committee then was the identification of anchor papers—exemplars that clearly and unambiguously represented the solid center of a score point as described in the scoring criteria. Those exemplary anchor papers formed the basis not only of reader training, but of subsequent range-finding discussions, as well.

Discussion was ongoing with the goal of identifying a sufficient pool of additional student responses for which consensus scores could be achieved and which illustrated the full range of student performance in response to the prompt or item. This pool of responses included borderline responses—ones that appeared to be between rather than clearly within a score level and which therefore represented a decision-making problem that readers (with training) would need to resolve.

TRAINING MATERIALS

Following the range-finding sessions, the performance assessment specialists and team leaders finalized the anchor sets and other training materials, as identified in the range-finding meetings. The final anchor papers were chosen for their clarity in exemplifying the criteria defined in the scoring rubrics.

The anchor set for each 4-point question consisted of a minimum of thirteen papers, three examples of each of the four score points and one example of a non-scorable paper. The anchor set for each 2-point question consisted of a minimum of seven papers, three examples of each of each score point and one example of a non-scorable paper. Score point exemplars consisted of one low, one solid mid-range, and one high example at each score point.

Additional training and qualifying sets of responses were selected to be used in reader training. One training set consisted of responses that were clear-cut examples of each score point; the second set consisted of responses closer to the borderline between two score points. The training sets gave readers an introduction to the variety of responses they would encounter while scoring, as well as allowing them to develop their decision-making capability for scoring responses that did not fall clearly into one of the scoring levels. Calibration/validity papers to be circulated during scoring were also identified at this time, as were reader qualifying sets.

RATER CONSISTENCY (RELIABILITY)

Reader training for each prompt was led by a performance assessment specialists and team leaders. The primary purpose of the training was to help the readers understand the decisions made by the range-finding committee. Also, training helped readers internalize the scoring rubrics, so that they might effectively and consistently apply them.

Reader training sessions included an introduction to the assessment itself. In addition, readers were informed of the parameters or context within which the students' performance was elicited. This gave readers a better understanding of what types of responses could be expected, given such parameters as grade level, instruction or time limitations. Readers next received a description of the scoring criteria that applied to the responses for each item.

The scoring criteria were always presented in conjunction with the anchor papers. After presentation and discussion of the anchor papers, each reader was given a training set consisting of ten papers. The readers scored the papers independently. When all readers had scored the training set, their preliminary scores were collected for reference.

Group discussion of the scores assigned was the next step, allowing the readers to raise questions about the application of the scoring rubric and giving them a context for those questions. The purpose of the discussion among the readers in training was to establish a consensus to ensure consistency of scores between readers. Even after readers had qualified for the scoring, training continued throughout the scoring of all responses to maintain high inter- and intra-reader reliability. Therefore, training was a continuous process and readers were consistently given feedback as they scored.

Frequent reliability checks were used to closely monitor the consistency of each reader's performance over time. The primary method of monitoring a reader's performance was by a process called "back-reading". In back-reading, each table leader reread and checked scores on an average of five to ten percent of each reader's work each day, with a higher percentage early in the scoring. If a reader was consistently assigning scores other than those the table leader would assign, the team leader and performance assessment specialist, together, retrained that reader, using the original anchor papers and training materials. This continuous, on-the-spot checking provided an effective guard against reader "drift," (beginning to score higher or lower than the anchor paper scores). Readers were replaced if they were unable to score consistently with the rubric and the anchor papers after significant training.

Tables 4-1 through 4-4 give the rater agreement information for the open-ended items in the 1999 Grade 10 WASL. Two types of rater agreement were calculated from 10 percent of the examinees randomly selected from the students' response booklets: score agreement for individual items and score agreement across the total score for the open-ended item set for each content area. For total score agreement on the open-ended items, the correlations were quite high (.96 to .98) within each content area with virtually no difference between the means of the total scores summed across open-ended items. For item-by-item interjudge agreement in Reading and Listening, the range of exact agreement was 74 to 97 percent and the range of exact and adjacent agreement was 99 to approximately 100 percent. For interjudge agreement in Writing, the range of exact agreement was 84 to 85 percent; exact and adjacent agreement was 99 to approximately 100 percent. For item-by-item interjudge agreement in Mathematics, the range of exact agreement was 70 to 91 percent and the range of exact and adjacent agreement was 97 to approximately 100 percent.

Table 4-1: 1999 Grade 10 Correlations between and Means of Total Scores of First and Second Readings for Open-Ended Items by Test.

Test	Correlation	Mean First Reading	Mean Second Reading
Listening & Reading	.97	16.24	16.03
Writing	.96	6.99	6.95
Mathematics	.98	16.85	16.85

Table 4-2: 1999 Grade 10 Frequencies of Exact Score Matches, Adjacent Scores, and Discrepant Scores for Listening and Reading Items.

Item	Points Possible	Exact Score Match	Adjacent Scores	Discrepant by Two Points	Discrepant by Three Points	Discrepant by Four Points	Percent Exact Agreement
7*	2	5859	1481	16			80%
8*	2	5513	1774	69			75%
2	2	6072	1275	9			83%
8	2	5706	1619	31			78%
13	2	7127	188	41			97%
15	2	5927	1380	49			81%
19	2	5879	1430	47			80%
22	2	7002	326	28			95%
24	4	5435	1838	80	2	1	74%
28	2	5813	1451	92			79%
30	4	5426	1821	99	9	1	74%
32	2	5843	1444	69			79%
35	2	5855	1449	52			80%
40	2	6540	798	18			89%

* Listening items

Table 4-3: 1999 Grade 10 Frequencies of Exact Score Matches, Adjacent Scores, and Discrepant Scores for Writing Scores.

Score	Points Possible	Exact Score Match	Adjacent Scores	Discrepant by Two Points	Discrepant by Three Points	Percent Exact Agreement
1	4	6596	1224	49	2	84%
2	2	6739	1122	30		85%
3	4	6694	1174	22	1	85%
4	2	6709	1160	22		85%

Table 4-4: 1999 Grade 10 Frequencies of Exact Score Matches, Adjacent Scores, and Discrepant Scores for Mathematics Items.

Item	Points Possible	Exact Score Match	Adjacent Scores	Discrepant by Two Points	Discrepant by Three Points	Discrepant by Four Points	Percent Exact Agreement
3	2	6653	642	20			91%
5	2	6326	945	44			86%
7	4	5793	1382	128	12		79%
11	2	5091	2126	98			70%
14	2	6416	875	24			88%
16	4	6102	916	243	37	17	83%
18	2	6003	1264	48			82%
21	2	6208	1092	15			85%
26	2	6642	640	33			91%
28	2	6428	869	18			88%
29	4	5535	1610	156	10	4	76%
33	2	5947	1332	36			81%
36	4	5296	1742	248	25	4	72%
37	2	6421	860	34			88%
40	2	6687	613	15			91%
42	2	6677	616	22			91%

Additional Considerations For Scoring Writing

Although the *training* for scoring writing is the same as described above, various approaches can be used in evaluation Writing. For the WASL, a "focused holistic" approach was selected. Focused holistic scoring, or general impression scoring, assesses relative writing fluency and measures the degree to which a writer has connected to the reader of a paper. When a paper is scored holistically, a reader considers the overall effectiveness of the piece of writing and assigns a score that reflects the reader's impression of the paper's overall quality. In a focused holistic approach, the reader also takes into account all of the elements that make up a successful piece of writing, for example content, organization, style, and mechanics. In the WASL Writing Test, Content, Organization, and Style are scored together on a 4-point scale and Writing Mechanics are scored on a 2-point scale. These two scores are combined to provide a maximum of 6 points on any one piece of writing.

PART 5

STANDARD SETTING PROCEDURES

Standard setting for the Grade 10 *Washington Assessment of Student Learning* (WASL) was conducted in the summer of 1999. Because all of the items in the WASL item pool are on the same underlying Rasch scale (see Part 2), these standards can be held consistent across different test forms, making it possible to monitor student achievement over time with a fixed performance standard in each content area.

Standard setting committees were composed of teachers, curriculum specialists in the relevant subject area, school administrators, parents, and community members (Table 5-1). All standard setting committee members had direct experience with fourth graders or with the curriculum materials relevant for fourth graders.

Table 5-1: Number of Standard Setting Judges in each Professional Role.

Professional Role	Number of Judges
Teachers	25
Specialist Teachers	2
School Administrator	4
Parent	5
Community Representative	3
Total	39

Setting standards for student performance on the WASL was essentially a systematic, judgmental process aimed at establishing a consensus, among knowledgeable people, about what fourth grade students should know and be able to do. Washington's Essential Academic Learning Requirements (EALRs) have described the expected content in Reading, Writing, Communications, and Mathematics for Washington's public schools (See Appendix A). The new assessments have defined, in performance terms, some of the important knowledge, skills, and abilities fourth grade students should demonstrate in relation to the EALRs. The purpose of the standard-setting process was to establish the level of performance expected of fourth grade students who are judged as meeting the standards in Listening, Reading, Writing, and Mathematics. The emphasis for the judges, in the standard setting process, was on what students should know and be able to do near the end of Grade 10.

Performance standards on the Grade 10 assessment were determined by the standard setting procedure described below. This procedure is particularly well adapted to setting standards on assessments with mixed item types (that is, multiple-choice, short-answer, and extended response formats) as used on WASL. The procedure used in Washington state has been applied successfully in other large-scale assessment programs and was reviewed and approved by the National Technical Advisory Committee (see Appendix E) for the Commission on Student Learning—a committee composed of nationally recognized measurement professionals.

READING, LISTENING, AND MATHEMATICS

Implementation of the standard setting process required that the judges first take the operational test just as the students experienced it. The judges also reviewed scoring guides for the constructed-response (short-answer and open-ended) items and examples of student responses anchoring each item's score points.

Next, each standard setting judge received a complete set of the items ordered by difficulty from easiest to hardest, rather than in the order they appeared in the students' test booklets. Multiple-choice items appeared only once in the ordered booklet. Two-and four-point items appeared two or four times, according to the difficulty of achieving each score point. Data from the spring 1997 operational assessment was used to establish item difficulties. The first item in the judges' ordered booklet was the easiest item on the test, that is, the one the highest number of students answered correctly. The last item in the judges' ordered booklet was the hardest item on the test, that is, the one the fewest number of students answered correctly. Although the judges knew the items were ordered from easiest to most difficult, they did not know how students actually performed on the items—that is, how many students answered item 1 correctly, item 2 correctly, and so forth.

In small groups, the judges examined the items in the ordered booklet one at a time, starting with the first (easiest) item in the booklet, and moving to the second easiest item, and so on, until all items (and their scoring rubrics) were examined. As judges examined each item, they were asked to consider:

- What is each item measuring?
- What makes each item more difficult than the items that precede it?

Judges proceeded through the ordered item booklets and trained table leaders encouraged them to observe the increase in the complexity of the items and note the increase in knowledge, skills, and abilities required to answer the items.

At the conclusion of this first review of the ordered booklets, judges were asked to make an individual decision about where to place a "flag" at "meets standard". Each flag was placed in the ordered item booklet according to the individual judge's expectation of what students who are performing at standard should know and be able to do. For example, each judge placed his or her "meets standard" flag at a location in the booklet such that if a student is able to respond correctly to the items that precede the flag (with at least 2/3 likelihood of success), then the student has demonstrated sufficient knowledge, skills, and abilities to infer that the student is performing at the standard. For multiple-choice items this means the student who "meets standard" should be likely to know the correct response. For short answer- or extended response-items (with multiple score points), this means the student who "meets standard" should be likely to achieve at least that score point.

For the Reading and Mathematics tests, judges were asked to insert two additional flags: one at "exceeds standard" and one between "near standard" (partially proficient) and "low" (minimal). In this way, progress toward or beyond standards could also be identified. These additional flags were not set for the Listening test because there were not a sufficient number of points on each test to warrant such a fine distinction of performance levels.

Because not all judges set their flags in the same locations, the next step involved each judge sharing and discussing the locations at which his or her flag(s) were placed. When one judge placed a flag for "meets standard" farther along in the ordered booklet than another judge, it implied that the first judge expected students who meet the standard to demonstrate a higher level of achievement on

the test. The difference in their individual expectations was reflected by the content and difficulty of the items between their flags.

For example, if Judge 1 placed a flag after item 30 and Judge 2 placed a flag after item 40, then these two judges disagreed on items 31-40. We know this because Judge 1, who placed a flag after item 30 was indicating that students who can correctly respond to the content in items 1-30 (with at least 2/3 likelihood) have demonstrated abilities sufficient to infer they have met the standard. Judge 2 (who placed the flag after item 40) did not agree, and was indicating that students have not demonstrated sufficient skills until they can handle more difficult content, that is, items 31-40.

Judges next discussed in small groups these differences in expectations as indicated by their different flag placements. Each group was provided with three lists indicating each judge's three flag locations for Reading and Mathematics. Beginning with the judges' placements of the "meets standard" flags, each judge was asked to note the location of every other judge's flag placement. Suppose the results in Table 5-2 occurred from the first round of standard setting.

Table 5-2: Example of Standard Setting Procedure

Judge Number	Meets Standard Flag Placed After:
1	item 30
2	item 34
3	item 29
4	item 33
5	item 36
6	item 39
7	item 33

Judges next would be asked to place a flag in their own ordered booklets after items 29, 30, 33, 34, 36, and 39. Now all judges could see the different expectations for student performance that "meet standard." In this example, judges would next discuss their differences, focusing on the items between 30 and 39 and discuss what these items ask of students' knowledge, skills, and abilities and whether students who meet the standard should be expected to respond correctly to these items. The discussion would consider the items one at a time beginning with item 30 and continuing up through item 39. When productive discussion of these items was completed, judges would then be asked to reevaluate their own initial flag locations in light of the small group discussion. Judges may decide to agree on a common flag placement during this round. That is, rather than requiring the calculation of the small group's average to determine the group's flag placement, the judges may agree to compromise and reach a consensus.

In the standard-setting for Reading and Mathematics, after judges had made their second round flag placements for "meets standard", the process was repeated for the other two cut-points—the below standard and the above standard locations.

Round 3 consisted of bringing the small groups back together as a large group to share and discuss each small group's flag placements. In the large group each judge placed a flag in his/her own ordered item booklet where each small group had made its flag placements. Large group discussion now focused on the items between the first and last flags for each performance level. Following the large group discussion, judges were asked to make a new (or reconfirm their former) flag placements.

Round 4 consisted of sharing with the large group the Round 3 small group results. Individual judges were then asked to make their final flag placements, which were then compiled to establish the final standard and other performance levels for each content area.

WRITING

Writing was handled in a slightly different manner than for Reading, Listening and Mathematics. There were two prompts (writing tasks). Each was scored for Content, Organization, and Style (1-4 points) and Mechanics (0-2 points). The scores from both prompts were combined (a possible range of 2-12 points) and the standard was set on the combined scores. To keep the standard-setting process for Writing as parallel with the other content areas as possible, the following standard-setting procedure was used:

1 Example responses were selected (both prompts together from the same student) that represented each of the possible combined score points 2-12 using a minimum of 3 students' responses for each possible score point.

2 These sets of combined student responses were ordered from lowest combined score (2) to highest combined score (12).

3 Judges were asked to proceed individually through all the example response sets (a minimum of 33) from lowest to highest and indicate the point at which the papers began to represent work "at the standard" and prior sets of papers represented work that was "less than the standard."

4 Next judges shared their individual judgments in their small groups and discussed the characteristics of the papers just above and just below their cut-points (flags).

5 The small group's placements were shared and discussed in the larger group.

6 Finally judges reconsidered their flags in light of the discussions and worked toward a consensus as to where the standard for Writing should be set.

SUMMARY

These processes ensured that the standards set for proficiency on the WASL tests would have careful scrutiny from a broad range of constituents of education. The judges had significant input from their peers and sufficient opportunities for discussion about their diverse opinions on standards.

PART 6
SCALE SCORES

All scaling for the Grade 10 *Washington Assessment of Student Learning* (WASL) was done using the same item data and calibrations used in the standard setting. Because the Mathematics and Reading Tests have four levels for student performance versus two levels for Listening and Writing, two different procedures were used to develop the scale scores. All four of the tests have a scale score of 400 representing the standard, but for Reading and Mathematics, the cut score for level two was set to equal 375 whereas in Listening and Writing an adjustment to the standard deviations was made to produce the scale scores. The following sections give details pertaining to the actual procedures used

DEVELOPMENT OF SCALE SCORES ON THE WASHINGTON ASSESSMENT OF STUDENT LEARNING

Scores on the WASL are reported as scale scores (See Tables 6-1 through 6-3 on Pages 8 through 10 of this chapter for 1999 Grade 10 number correct to scale scores conversions for each test.). As described in Part 2, the Rasch model and Master's (1982) extension of the Rasch model to multiple point items (the partial credit model) result in an equal interval scale (much like a ruler that is marked in inches or centimeters) for each test on which items and student scores can be reported. The partial credit model (PCM) allows for the inclusion of open-ended items where the maximum points possible are greater than one. Calibrating a test with Master's partial credit model produces estimated item parameters for an item's difficulty and the difficulty of its various score points (or steps). The possible scale score range for the WASL across the four test scales is 100 to 650 given all of the items in the item pool. This range is sufficient to describe levels of performance from the lowest possible earned scale score to the highest possible earned scale score *across all content areas tested and across different test forms*. The actual range of scale scores each year and in each content area will differ. For example, the range of possible scale scores for the Grade 10 1998 Mathematics Test is 195 to 580 (See Table 6-3).

The Rasch model is an item response theory (IRT) model. IRT models can generate three parameters for items: item difficulties, item discriminations, and guess levels (the probability that low achieving examinees can guess correctly on multiple-choice items). The Rasch and PCM models also generate theta (θ) for each examinee. Because Rasch models treat all items as equally discriminating and assume that there is no guessing, there are no item discrimination and guessing parameters calculated. This means that, unlike more complicated scoring models, there is a one to one relationship between the number correct score on a test and the θ score on the test.

Once θ scores are generated, it is general practice to convert θ to a positive, whole number scale through a linear conversion procedure. The resulting numbers on the whole number scale are easy to use for computations when generating district, school, or building averages.

Because the scaled scores are on an equal interval scale, it is possible to compare score performance at different points on the scale. Much like a yard-stick, differences are constant at different measurement points. For example, a difference of 2 inches between 12 and 14 inches is the same differences as a difference of 2 inches between 30 and 32 inches. Two inches is two inches. Similarly, for equal interval achievement scales, a difference of 40 scaled score points between 360 and 380 means the same difference in achievement as a difference of 400 and 420, except that the difference is in degree of achievement rather than length.

The major limitation of scaled scores is that they are not well suited to making score interpretations beyond "how much more" and "how much less". Administrators, parents, and students ask, "What score is good enough? How do we compare with other schools like ours? Is a 40 point difference between our school and another school a meaningful difference?" For this reason, scale scores are usually interpreted by using performance standards or by converting them to percentile ranks.

Based on the content of the WASL, committees set the performance standards for each test (Reading, Writing, Listening, and Mathematics) that would represent acceptable performance for a well taught, hard working seventh grade student (see Part 4). In Reading and Mathematics, the standard setting committees also identified two "below standard" and one "above standard" performance levels[2]. Because the Listening and Writing Tests were relatively short, only two performance levels were established - "meets standard" and "does not meet standard."

The standard setting (described in Part 4) allowed the standard setting committees to identify the θ values associated with each cut-score (i.e., in Reading and Mathematics, the cut between "substantially below standard" and "approaches standard", between approaches standard and "meets standard", and finally between "meets standard" and "exceeds standard"; in Writing and Listening, the cut between "does not meet standard" and "meets standard"). It was these θ values that formed the basics for the scaling procedure. In order to maintain the linear scale defined by the raw score to θ relationship, any two points on the θ scale can be fixed to scale scores and the resulting transformation will remain linear. That is what was done here.

Reading and Mathematics

Following the standard setting process, a linear conversion was used to transform the θ (logistic ability) scores (from the PCM analyses) to a whole number scale. For all tests, the θ score identified as "meets standard" was converted to a WASL scale score of 400. For Reading and Mathematics, the θ score identified as "below standard-level 2" was converted to a Washington scale score of 375. The rest of the θ scores were converted to the whole number scale using the linear conversion equations for each test that produced these two scale score points. Only two points can be set in a linear transformation and all other points must be derived from the conversion formula.

---

[2] The following are the general descriptions of the performance levels established for the Washington Assessment of Student Learning:

Level 4 -- Above Standard: This level represents superior performance, notably above that required for meeting the standard at grade 10.

Level 3 -- MEETS STANDARD: This level represents solid academic performance for grade 10. Students reaching this level have demonstrated proficiency over challenging content, including subject-matter knowledge, application of such knowledge to real world situations, and analytical skills appropriate for the content and grade level.

Level 2 -- Below Standard: This level denotes partial accomplishment of the knowledge and skills that are fundamental for meeting the standard at grade 10.

Level 1 -- Well Below Standard: This level denotes little or no demonstration of the prerequisite knowledge and skills that are fundamental for meeting the standard at grade 10.

In all content areas, the standard (Level 3) reflects what a well taught, hard working student should know and be able to do.

Therefore, the "above standard" scale score for Reading was set at 416 and the "above standard" scale score for Mathematics was set at 432.

The general formula for a linear equation converting θ to a scaled score is:

$$\theta a + b = \text{scaled score} \qquad (6\text{-}1)$$

Where **a** is a distribution variable for the whole number scaled scores and **b** is a location on the whole number scale.

To obtain the linear formula necessary to translate from the θ scale to the whole number scale for Reading and Mathematics, the scaled score cut points for "meets standard" (400) and approaches standard (375) are plugged into the above formula and, through simultaneous solution of two equations, one can solve for **a** and **b.**

For math, the point on the θ scale where the standard setting committee decided that students had "met standard" was 0.286 and the point on the θ scale where the standard setting committee decided that students were "approaching standard" was -0.349. Therefore the initial linear equations were:

$$0.286\mathbf{a} + \mathbf{b} = 400 \qquad (6\text{-}2)$$
$$-0.349\mathbf{a} + \mathbf{b} = 375 \qquad (6\text{-}3)$$

Solving for a and b, the results are **a** = 39.37 and **b** = 388.74. These values were then used with the Mathematics θ scores to transform all θ scores to Mathematics scaled scores.

$$\text{Mathematics Scaled Score} = 39.37(\theta) + 388.74 \qquad (6\text{-}4)$$

For Reading, the point on the θ scale where the standard-setting committee decided that students had "met standard" was 0.793 and the point on the θ scale where the standard setting committee decided that students were "approaching standard" was –0.110. Therefore the initial linear equations were:

$$0.793\mathbf{a} + \mathbf{b} = 400 \quad (6\text{-}5)$$
$$-0.110\mathbf{a} + \mathbf{b} = 375 \quad (6\text{-}6)$$

Solving for a and b, the results are **a** = 27.69 and **b** = 378.05. These values were then used with the Reading θ scores to transform all θ scores to Reading scaled scores.

$$\text{Reading Scaled Score} = 27.69(\theta) + 378.04 \quad (6\text{-}7)$$

In Reading and Mathematics, students who earn scale scores below 375 are placed in the "below standard-level." Students who earn scale scores of 375 to 399 are placed in the "below standard, level 2" category in both Reading and Mathematics. Students who earn scale scores of 400 to 413 in Reading or 400 to 426 in Mathematics are in the "meets standard" category. Students who earn scale scores of 414 and higher in Reading or 427 and higher in Mathematics are in the "above standard" category.

Listening and Writing

In the standard setting for Listening and Writing only a single cut score was set representing the standard. Therefore the linear transformations θ for Listening and Writing required that one additional point be set. The decision was made to set the standard deviations of the θ scale of each test to a value so that the range of scale scores was within a 100 to 650 range obtained for the Reading and Mathematics Tests. Once the linear transformation formula was obtained, all θ for the Listening and Writing Tests were converted to whole number scaled scores. This means that scale scores of 400 or higher meet the standard in all content areas and scale scores of 399 or lower are below the standard.

CUT POINTS FOR CONTENT STRANDS

The cut points for the individual *content strands* in Reading and Mathematics were determined in the following manner. Using the θ value associated with "meets standard" and the item difficulties, it was possible to estimate the score of a proficient examinee on each of the items within the strand. Figure 6-1 gives a hypothetical distribution of item difficulties for the items in the Mathematics strands. As can be seen, the range of item difficulties differs for each strand. What may be less apparent is that the number of items below and above the theta value of .286 also differs. Students receiving raw scores for each of the strands equal to or higher than the estimated strand score for proficient examinees are reported as "similar to the performance expected of students who met the standard". Raw scores below this cut point are reported as "below the performance expected of students who met the standard". In Listening there are no scores reported at the strand level.

The Writing Test consists of only two writing prompts, so using the partial credit model is not appropriate. Instead all scaling was done on the raw score scale. In Writing the

Figure 6-1: Hypothetical Range of Item Difficulties (theta values) within Mathematics Strands

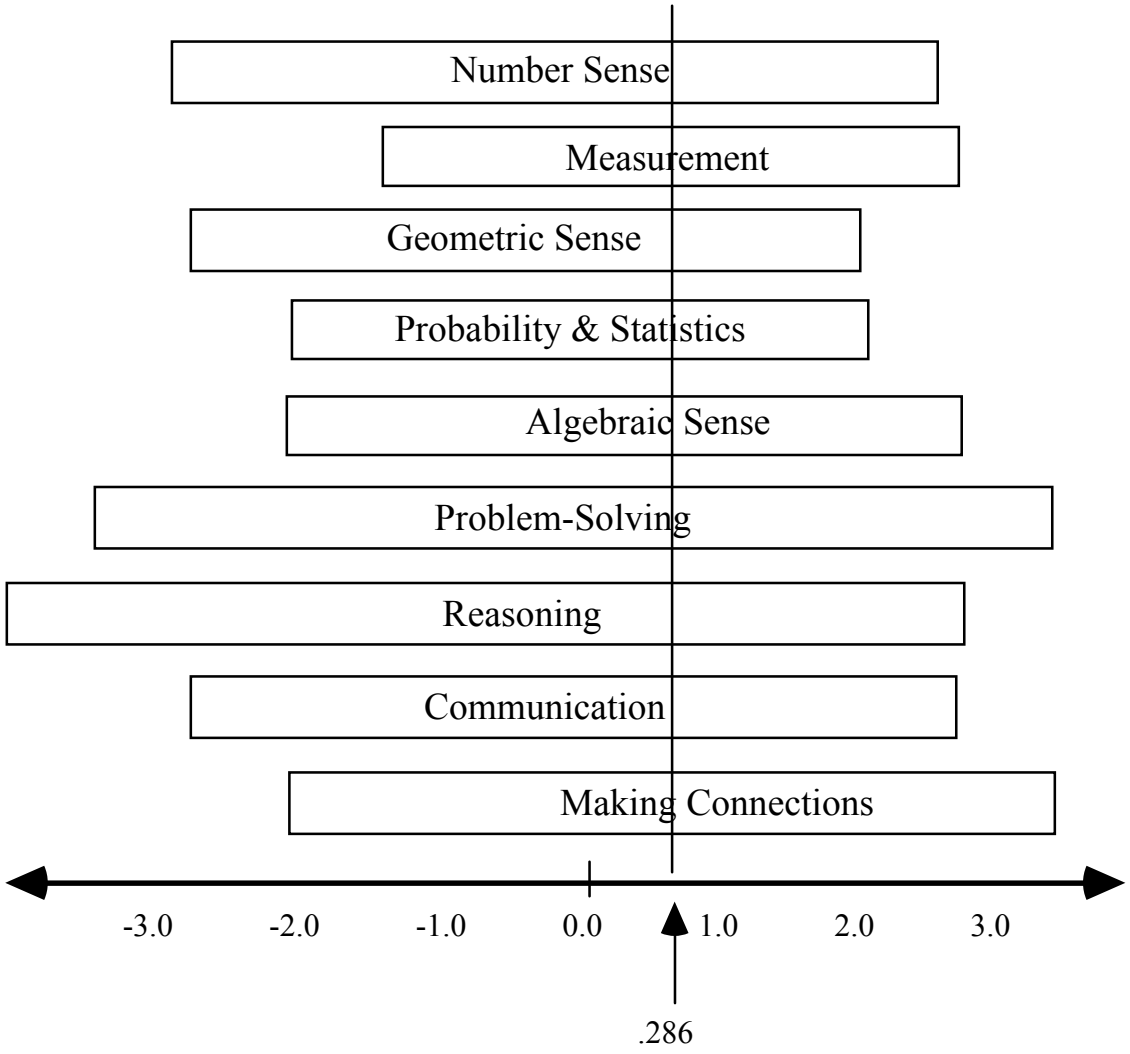
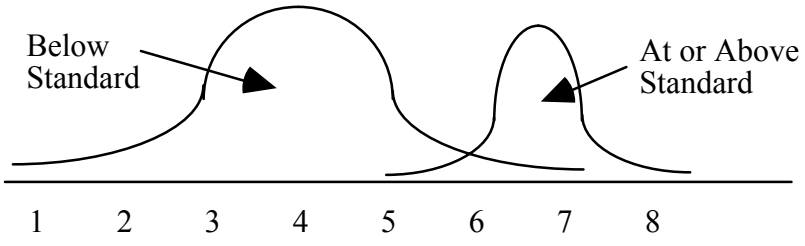


Figure 6-2: Score Distribution of Students Identified as Below Standard and Score Distribution of Students Identified to Be At or Above Standard: Content, Organization, and Style



cut-score for the two strands were determined in the following manner. The data from the standard-setting was divided into two sets, one consisting of examinees meeting the standard, the other examinees not meeting the standard. The raw scores for Writing Content, Organization, and Style and for Writing Mechanics were obtained for the examinees in each group (those meeting the standard and those not meeting the standard). Frequency distributions were computed on each of the strands for each group. Cut-points were identified as those showing the smallest overlap between the distributions of the two groups (see Figure 6-2). This is often referred to as a "contrasting groups design". Discussions of the standard setting committees also contributed to the decision. In the end, a minimum combined score of six for the Writing Content, Organization, and Style strand and a combined score of three for the Writing Mechanics strand were determined to be the cut points and the item parameters.

EQUATING

The score scales established for the Grade 10 WASL in 1999 will stay in place for all subsequent years and test forms unless the scale is changed and new standards are set. Although new test forms are developed each year, Listening, Reading, and Mathematics will be equated using calibrations to items that were used in the base operational year (1999) – thus maintaining the same scale score system, i.e., 400 for meeting the standard. Although the raw score to scale score relationship will change for Listening, Reading, and Mathematics, the level of difficulty associated with meeting the standard in each tested content area will remain statistically equivalent over time.

NUMBER CORRECT SCORES TO SCALE SCORES

Each year WASL tests will have a different number correct score (raw score) to scale score relationship, although the underlying scale remains the same from year to year. This is possible because all items in the pool are on the same underlying Rasch scale. Table 6-1 gives the number correct score (NCS) to scale score (SS) relationship for the 1999 Grade 10 WASL Listening Test. Table 6-2 gives the NCS to SS relationship for the 1999 Grade 10 Reading Test. Table 6-3 gives the NCS to SS relationship for the 1999 Grade 10 Reading Test.

Table 6-1: 1999 Grade 10 Listening Number Correct Scores (NCS) to Scale Scores (SS)

NCS	Listening SS
0	203
1	237
2	276
3	305
4	331
5	356
6	381
7	408
8	437
9	476
10	509

Table 6-1: 1999 Grade 10 Reading Number Correct Scores (NCS) to Scale Scores (SS)

NCS	Reading SS
0	243
1	262
2	283
3	295
4	304
5	311
6	318
7	323
8	328
9	332
10	336
11	340
12	343
13	347
14	350
15	353
16	356
17	358
18	361
19	364
20	366
21	369
22	371
23	373
24	376
25	378
26	380
27	382
28	385

NCS	Reading SS
29	387
30	389
31	391
32	393
33	395
34	398
35	400
36	402
37	404
38	407
39	409
40	411
41	414
42	416
43	419
44	422
45	425
46	428
47	431
48	435
49	439
50	443
51	448
52	454
53	462
54	473
55	492
56	511

Table 6-2: 1999 Grade 10 Mathematics Number Correct Scores (NCS) to Scale Scores (SS)

NCS	Mathematics SS	NCS	Mathematics SS
0	195	36	391
1	223	37	393
2	251	38	395
3	268	39	397
4	280	40	400
5	289	41	401
6	297	42	403
7	304	43	405
8	310	44	408
9	315	45	410
10	320	46	412
11	324	47	414
12	329	48	417
13	332	49	419
14	336	50	422
15	339	51	424
16	343	52	427
17	346	53	430
18	349	54	433
19	352	55	436
20	354	56	439
21	357	57	443
22	360	58	447
23	362	59	451
24	365	60	455
25	367	61	460
26	369	62	465
27	372	63	471
28	374	64	478
29	376	65	486
30	378	66	496
31	380	67	508
32	383	68	525
33	385	69	553
34	387	70	580
35	389		

Reference

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrica*, (*47*), 149-174.

PART 7

RELIABILITY

The reliability of test scores is a measure of the degree to which the scores on the test are a "true" measure of the examinees' knowledge and skill relevant to the tested knowledge and skills. Simply put, the reliability is the proportion of observed score variance that is true score variance.

There are several ways to obtain estimates of score reliability: test-retest, alternate forms, internal consistency, and generalizability analysis are the most common. Test-retest estimates require administration of the same test at two different times. Typically the testing times for achievement tests are close together so that new learning does not impact scores. Alternate forms reliability estimates require administration of two parallel tests. These tests must be created in such a way that we have confidence that they measure the same domain of knowledge and skills using different items. Both test-retest and alternate forms estimates of the reliability of scores require significant testing time for examinees and are generally avoided when there is a concern that fatigue or loss of motivation might impact the resulting reliability coefficient.

The *Washington Assessment of Student Learning* (WASL) is a rigorous measure that requires significant concentration on the part of students for a sustained period of time. For this reason, it was determined that test-retest and alternate forms reliability methods were unlikely to yield accurate estimates of score reliability. Therefore, internal consistency measures were used to estimate score reliability for Reading, Listening, Writing, and Mathematics tests.

INTERNAL CONSISTENCY

Internal consistency reliability is an indication of how similarly students perform across items measuring the same knowledge and skills—in other words, how consistent each examinee performs across all of the items within a test. Internal consistency can be estimated using Cronbach's alpha coefficient. When a test is composed entirely of multiple-choice (dichotomously scored) items, a modification of Cronbach's alpha can be used (KR-20). However, when multiple-point items are included on a test, Cronbach's alpha coefficient provides the internal consistency estimate. Two of the demands for applying this method when estimating score reliability are: 1) the number of items should be sufficient to obtain stable estimates of students' achievement and 2) all test items should be homogeneous (similar in format and measuring very similar knowledge and skills).

WASL Reading and Mathematics tests have sufficient items to address the issue of test length; however, the Listening Test has fewer items/scores, hence this will have a tendency to depress the alpha coefficient. In addition, the Listening Test scores are generally high with a mean of 7.9 out of 10 possible points. This may also depress the alpha coefficient due to a restriction in the range of scores.

WASL is also a complex measure that combines multiple-choice, short-answer, and extended response items. The Mathematics and Reading tests measure multiple strands that are all components of the domains of Mathematics and Reading respectively. Hence, examinee performance may differ markedly from one item to another due to prior knowledge, educational experiences, exposure to similar content, etc. Because of the heterogeneity of items in the Reading and Mathematics tests and the short test length for the Listening test, use of Cronbach's alpha for estimating score reliability for WASL will likely *under-estimate* of the actual reliability of scores. When items are heterogeneous,

as they are in the WASL, it is generally believed that the true score reliability is higher than the estimate obtained through the alpha coefficient.

The WASL Writing Test is composed of two written essays. Although there are only four scores for the test (two for each of the essays), the items measure essentially the same ability twice. These items are very homogeneous; therefore, the alpha coefficient may be a reasonable estimate of the reliability of the scores.

The alpha coefficient is obtained through the following formula:

$$r_{xx'} = \left[\frac{N}{N-1}\right]\left(1 - \frac{\sum s_i^2}{s_x^2}\right)$$

Where:

$\sum s_i^2$ is the sum of all of the item variances

$\sum s_x^2$ is the observed score variance, and

N = the number of items on the test

Alpha coefficients for each of the 1999 Grade 10 WASL tests are given in Table 7-1. As can be seen, scores from the longer tests have higher reliability estimates. However, even with the very short Listening and Writing tests, these estimates provide good evidence for the overall reliability of 1999 Grade 10 WASL test scores.

Table 7-1: 1999 Grade 10 Reliability Estimates and Standard Error Of Measurement for Each WASL Test

Subtest	Alpha Coefficient	Scaled Score† or Raw Score Standard Error* of Measurement
Listening†	.77	27.6
Reading†	.92	8.4
Mathematics†	.93	11.4
Writing*	.85	1.0

STANDARD ERROR OF MEASUREMENT

One way to interpret the reliability of test scores is through the use of the Standard Error of Measurement (s_{em}). The s_{em} is an estimate of the standardized distribution of error around a given observed score. When one s_{em} is added and subtracted from an observed score, we can be about 68 percent certain that the examinee's true score lies within the band. For example, the s_{em} for the 1999 Grade 10 Reading Test is 8.4. If the examinee's scale score was 402, we could be about 68 percent certain that the examinee's true score was between 402 – 8.4 and 402 + 8.4 or between 393.6 and 410.4. If we add and subtract two s_{em}, we can be about 95 percent certain that the examinee's true

score lies between 385.2 and 418.8. Finally, if we add and subtract three s_{em}, we can be about 99 percent certain that the examinee's true score lies between 376.8 and 427.2. In classical testing, we obtain the s_{em} through the following formula:

$$s_{em} = s_x \sqrt{1 - r_{xx'}}$$

Where:

s_x is the observed score standard deviation, and

$r_{xx'}$ is the reliability estimate (alpha)

Table 7-1 provides the 1999 Grade 10 standard error of measurement for the scaled scores of WASL Reading, Listening and Mathematics Tests based on the standard deviation of the scale scores and the alpha coefficient. Table 7-1 also gives the 1999 Grade 10 standard error of measurement for the raw scores of the WASL Writing Test based on the standard deviation of the raw scores and the alpha coefficient.

INTERJUDGE AGREEMENT

As was described in Part 4, inter-judge (inter-rater) agreement was another important source of evidence for the reliability of test scores. When two trained judges agree with the score given to a student's work, this gives support for the score on the short-answer or extended response item. Two methods are described in Part 4 for determining the degree to which judges gave equivalent score to the same student work: correlations between totals, when scores for open-ended items are summed, and percent agreement. For total score agreement on the open-ended items, the correlations were .96 to .98 across content areas with virtually no difference between the means of the total scores summed across open-ended items. For item-by-item interjudge agreement in Reading and Listening, the range of exact agreement was 74 to 97 percent and the range of exact and adjacent agreement was 99 to approximately 100 percent. For interjudge agreement in Writing, the range of exact agreement was 84 to 85 percent; exact and adjacent agreement was 99 to approximately 100 percent. For item-by-item interjudge agreement in Mathematics, the range of exact agreement was 70 to 91 percent and the range of exact and adjacent agreement was 97 to approximately 100 percent.

SUMMARY

In summary, the data from the interjudge agreement study indicates that the judges can consistently score performances using the scoring criteria developed for each item. Data from the alpha coefficients indicate that, except for the listening test, the test scores can be trusted to represent examinees' performance on the concepts and skills measured by the test. Standard errors of measurement, however, are large enough that caution should be used when evaluating and making decisions based on individual students' scores.

PART 8

DESCRIPTION OF PERFORMANCE FOR 1999 GRADE 10 STUDENTS

The data presented in this section of the report is descriptive of performance of fourth grade students throughout the state on the 1999 *Washington Assessment of Student Learning* (WASL). Included are means, standard deviations, and numbers tested for the all tested fourth graders and disaggregated by a variety of groups (Tables 8-1 through 8-14). Also presented are the percent of students in each gender, ethnic, and categorical program group who met or did not meet the standards for each content area (Tables 8-15 through 8-26). These data are useful for tracking, over time, the state's progress in helping students meet the Essential Academic Learning Requirements. *One possible limitation to the data is that the categorization of students is based on the way students are identified on their response books. If response books for given students did not indicate gender, ethnicity, and/or categorical program, the data for these students are not included in disaggregated data.* Finally, Tables 8-27 through 8-30 provide the mean performance on each item of the Grade 10 WASL tests, as well as the item-test correlations for each item.

SUMMARY STATISTICS

The means for each score were computed by summing the relevant scores for all students tested and dividing by the total number of students tested. The standard deviation was computed by obtaining the square root of the relevant variances using the following equation:

$$SD = \sqrt{\frac{\Sigma\left(X - \overline{X}\right)^2}{N}}$$

where:

X is the individual score

$\overline{X}$ is the mean of scores for all students tested in the state, and

N is the number of students tested in the state (those with valid scores)

Table 8-1 provides the state summary statistics for those Grade 10 students taking the WASL tests in 1999. The column headed "Points Possible" contains the maximum number of scale score points possible in each test for the 1999 form. The next two columns contain the mean scale score and standard deviation of the scale scores for all students tested in the state. Table 24 provides the state 1999 Grade 10 summary statistics for the WASL strands within tests. The column headed "Points Possible" indicates the maximum number of points possible in each strand for the 1999 form. The next two columns contain the mean number correct strand score and standard deviation of the strand scores for all students tested in the state. The final column indicates the percent of students whose performance on the strand was similar to those who met the standard. Tables 25 through 28 provide the summary data for each ethnic and gender group tested in 1999 (as recorded on the response books). Table 29 through 32 provide the summary data for students in each of the following categorical programs: Learning Assistance Program (LAP) Reading, LAP Mathematics, Title 1 Reading, Title 1 Mathematics, Title 1 School, Bilingual/English as a Second Language (ESL), Highly Capable Students, Section 504, Special Education, and Migrant Education (as recorded on the response books).

Table 8-1: 1999 Grade 10 Scale Score Means, Standard Deviations, and Maximum Scale Scores by Test

Test	Number Tested	Maximum Scale Score† or Raw Score*	Mean Scale Score or Raw Score	Standard Deviation
Listening†	64138	509	441.08	57.38
Reading†	63040	511	402.78	29.54
Writing*	60742	12	8.18	2.55
Mathematics†	65270	580	382.23	42.83

Table 8-2: 1999 Grade 10 Maximum Number Possible, Number Correct Score Means, Standard Deviations (SD) by Strand, and Percent of Students with Strength in Strand

Strand	Number with Valid Scores	Points Possible	Mean	SD	Percent with Strength in Strand
Main Ideas & Details of Fiction	63040	6	4.28	1.44	43.7
Analysis and Interpretation of Fiction	63040	9	6.14	1.89	58.8
Thinks Critically about Fiction	63040	12	6.34	3.07	52.1
Main Ideas & Details of Nonfiction	63040	8	5.80	1.79	54.8
Analysis and Interpretation of Nonfiction	63040	11	6.53	2.63	47.6
Thinks Critically about Nonfiction	63040	10	6.38	2.63	47.0
Writing Content, Organization Style	60742	8	5.42	1.61	43.7
Writing Mechanics	60742	4	2.76	1.23	47.4
Number Sense	65270	8	4.29	1.92	27.4
Measurement	65270	7	3.77	1.86	34.7
Geometric Sense	65270	8	3.36	1.99	27.2
Probability & Statistics	65270	8	4.69	2.16	36.6
Algebraic Sense	65270	8	3.00	2.18	33.9
Solves Problems	65270	9	3.56	2.94	34.0
Reasons Logically	65270	8	3.57	2.32	32.2
Communicates Understanding	65270	8	3.97	2.24	39.9
Makes Connections	65270	6	2.61	1.83	28.6

Table 8-3: 1999 Grade 10 Listening Test: Number Tested, Scale Score Means, and Standard Deviations (SD) by Gender

Gender	Number Tested	Mean	SD
Females	31436	448.32	53.88
Males	32436	434.25	59.67

Table 8-4: 1999 Grade 10 Listening Test: Number Tested, Scale Score Means, and Standard Deviations (SD) by Ethnic Group

Ethnic Group	Number Tested	Mean	SD
African American/Black	2364	414.37	64.47
Alaska Native/Native American	1262	413.29	65.27
Asian/Pacific Islander	4639	432.30	63.46
Latino/Hispanic	4145	403.03	66.06
White/Caucasian	48308	447.61	53.07
Multi-Ethnic	2504	438.82	56.68

Table 8-5: 1999 Grade 10 Reading Test: Number Tested, Scale Score Means, and Standard Deviations (SD) by Gender

Gender	Number Tested	Mean	SD
Females	30941	406.86	28.20
Males	31855	398.93	30.23

Table 8-6: 1999 Grade 10 Reading Test: Number Tested, Scale Score Means, and Standard Deviations (SD) by Ethnic Group

Ethnic Group	Number Tested	Mean	SD
African American/Black	2223	384.64	28.96
Alaska Native/Native American	1216	388.07	29.07
Asian/Pacific Islander	4586	398.77	29.90
Latino/Hispanic	4018	383.10	29.43
White/Caucasian	47679	406.35	28.40
Multi-Ethnic	2449	399.95	28.02

Table 8-7: 1999 Grade 10 Writing Test: Number Tested, Raw Score Means, and Standard Deviations (SD) by Gender

Gender	Number Tested	Mean	SD
Females	30268	8.78	2.33
Males	30256	7.59	2.61

Table 8-8: 1999 Grade 10 Writing Test: Number Tested, Raw Score Means, and Standard Deviations (SD) by Ethnic Group

Gender or Ethnic Group	Number Tested	Mean	SD
African American/Black	2077	7.01	2.58
Alaska Native/Native American	1111	7.10	2.55
Asian/Pacific Islander	4413	8.32	2.60
Latino/Hispanic	3816	6.62	2.69
White/Caucasian	46209	8.41	2.46
Multi-Ethnic	2324	7.89	2.57

Table 8-9: 1999 Grade 10 Mathematics Test: Number Tested, Scale Score Means, and Standard Deviations (SD) by Gender

Gender	Number Tested	Mean	SD
Females	31910	381.64	40.77
Males	33080	383.02	44.67

Table 8-10: 1999 Grade 10 Mathematics Test: Number Tested, Scale Score Means, and Standard Deviations (SD) by Ethnic Group

Ethnic Group	Number Tested	Mean	SD
African American/Black	2417	350.45	37.36
Alaska Native/Native American	1300	358.50	41.33
Asian/Pacific Islander	4707	386.67	43.25
Latino/Hispanic	4250	353.48	38.90
White/Caucasian	49121	387.19	41.48
Multi-Ethnic	2545	374.59	41.11

Table 8-11: 1999 Grade 10 Listening Test: Number Tested, Scale Score Means, and Standard Deviations (SD) by Categorical Program

Categorical Program	Number Tested	Mean	SD
LAP Reading	75	404.55	58.96
LAP Mathematics	39	389.41	59.28
Title 1 Reading	576	397.73	66.38
Title 1 Mathematics	495	394.24	68.05
Section 504	338	427.66	56.80
Special Education	4012	380.27	65.26
Title 1 Migrant Education	341	372.94	63.78
Bilingual/ESL	1620	358.46	62.41
Gifted/Highly Capable Students	1067	473.08	41.47

Table 8-12: 1999 Grade 10 Reading Test: Number Tested, Scale Score Means, and Standard Deviations (SD) by Categorical Program

Categorical Program	Number Tested	Mean	SD
LAP Reading	76	377.29	24.69
LAP Mathematics	40	367.80	30.65
Title 1 Reading	574	379.60	27.94
Title 1 Mathematics	483	379.98	29.31
Section 504	332	392.04	26.98
Special Education	3840	367.42	29.12
Title 1 Migrant Education	331	371.66	28.30
Bilingual/ESL	1561	362.78	26.54
Gifted/Highly Capable Students	1037	428.69	26.65

Table 8-13: 1999 Grade 10 Writing Test: Number Tested, Raw Score Means, and Standard Deviations (SD) by Categorical Program

Categorical Program	Number Tested	Mean	SD
LAP Reading	67	6.76	2.71
LAP Mathematics	33	5.64	2.71
Title 1 Reading	524	6.51	2.72
Title 1 Mathematics	442	6.52	2.81
Section 504	307	7.16	2.50
Special Education	3463	5.09	2.42
Title 1 Migrant Education	317	5.58	2.54
Bilingual/ESL	1380	5.26	2.41
Gifted/Highly Capable Students	1021	10.12	1.94

Table 8-14: 1999 Grade 10 Mathematics Test: Number Tested, Scale Score Means, and Standard Deviations (SD) by Categorical

Categorical Program	Number Tested	Mean	SD
LAP Reading	78	339.13	31.99
LAP Mathematics	44	335.64	33.37
Title 1 Reading	604	345.24	35.85
Title 1 Mathematics	522	344.76	35.69
Section 504	347	369.10	40.76
Special Education	4164	333.35	38.16
Title 1 Migrant Education	348	338.82	34.76
Bilingual/ESL	1683	343.59	37.93
Gifted/Highly Capable Students	1066	427.06	37.79

PERCENT MEETING STANDARD

Tables 8-15 through 8-22 provide the 1999 information regarding the number of students in each gender and ethnic group (as indicated on the response books) who met the standard in Listening, Reading, Writing, and Mathematics. Tables 23 through 30 provide the information regarding the number of students in each categorical program (as indicated on the response books) who met the standard in Listening, Reading, Writing, and Mathematics in 1999. The following are the general descriptions of the performance levels established for the Washington Assessment of Student Learning:

Level 4 Above Standard: This level represents superior performance, notably above that required for meeting the standard at grade 10.

Level 3 MEETS STANDARD*: This level represents solid academic performance for grade 10. Students reaching this level have demonstrated proficiency over challenging content, including subject-matter knowledge, application of such knowledge to real world situations, and analytical skills appropriate for the content and grade level.

Level 2 Below Standard: This level denotes partial accomplishment of the knowledge and skills that are fundamental for meeting the standard at grade 10.

Level 1 Well Below Standard: This level denotes little or no demonstration of the prerequisite knowledge and skills that are fundamental for meeting the standard at grade 10.

** In all content areas, "Meets Standard" reflects what a well taught, hard working student should know and be able to do.*

For the Writing and Listening Tests, the tables show, for each group, the percent meeting standard and the percent not meeting standard. For the Reading and Mathematics tests, the tables show, for each group, the percent in each performance level. For Reading and Mathematics, students in Levels 1 and 2 did not meet the standard. Students in Levels 3 and 4 met or exceeded the standard.

Table 8-15: 1999 Grade 10 Listening Test: Percent Meeting Standards by Total (N=72,279) and by Gender

Group	Percent Meeting Standard	Percent Not Meeting Standard	Percent Not Tested	Percent Exempt
All Students	72.6	13.9	10.2	3.3
Females	76.2	14.1	6.8	2.8
Males	68.0	20.0	8.1	3.8

Table 8-16: 1999 Grade 10 Listening Test: Percent Meeting Standards by Ethnic Group

Ethnic Group	Number of Students	Percent Meeting Standard	Percent Not Meeting Standard	Percent Not Tested	Percent Exempt
African American/Black	2938	51.5	29.0	12.6	6.9
Alaska Native/Native American	1557	53.1	28.0	13.6	5.3
Asian/Pacific Islander	5155	66.4	23.6	6.8	3.2
Latino/Hispanic	5025	46.3	36.1	11.3	6.2
White/Caucasian	53449	76.8	13.6	6.7	2.9
Multi-Racial	2665	75.3	18.7	5.6	0.4

Table 8-17: 1999 Grade 10 Reading Test: Percent Meeting Standards by Total (N=72,279) and by Gender

	Meets Standard		Does Not Meet Standard			
Group	Percent Level 4	Percent Level 3	Percent Level 2	Percent Level 1	Percent Not Tested	Percent Exempt
All Students	33.4	18.1	23.1	13.7	11.7	3.3
Females	37.8	18.7	22.1	10.3	8.3	2.8
Males	28.8	17.2	23.8	16.6	9.8	3.8

Table 8-18: 1999 Grade 10 Reading Test: Percent Meeting Standards by Ethnic Group

		Meets Standard		Does Not Meet Standard			
Ethnic Group	Number of Students	Percent Level 4	Percent Level 3	Percent Level 2	Percent Level 1	Percent Not Tested	Percent Exempt
African American/Black	2938	11.3	12.9	26.4	25.0	17.4	6.9
Alaska Native/Native American	1557	14.1	14.0	28.0	22.0	17.0	4.9
Asian/Pacific Islander	5155	29.2	17.7	24.0	18.1	8.0	3.1
Latino/Hispanic	5025	11.8	12.5	26.0	29.7	13.9	6.1
White/Caucasian	53449	37.6	18.9	22.1	10.6	7.9	2.9
Multi-Racial	2665	30.6	19.6	25.9	15.8	7.7	0.4

Table 8-19: 1999 Grade 10 Writing Test: Percent Meeting Standards by Total (N=72,279) and by Gender

Group	Percent Meeting Standard	Percent Not Meeting Standard	Percent Not Tested	Percent Exempt
All Students	41.1	40.8	14.6	3.6
Females	50.3	36.7	10.0	3.0
Males	31.6	50.5	13.8	4.1

Table 8-20: 1999 Grade 10 Writing Test: Percent Meeting Standards by Ethnic Group

Ethnic Group	Number of Students	Percent Meeting Standard	Percent Not Meeting Standard	Percent Not Tested	Percent Exempt
African American/Black	2938	20.7	50.0	21.9	7.4
Alaska Native/Native American	1557	21.2	50.2	22.7	5.9
Asian/Pacific Islander	5155	43.1	42.5	11.1	3.3
Latino/Hispanic	5025	19.3	56.6	17.4	6.6
White/Caucasian	53449	44.6	41.9	10.4	3.1
Multi-Racial	2665	37.4	49.8	12.3	0.5

Table 8-21: 1999 Grade 10 Mathematics Test: Percent Meeting Standards by Total (N=72,279) and Gender

	Meets Standard		Does Not Meet Standard			
Group	Percent Level 4	Percent Level 3	Percent Level 2	Percent Level 1	Percent Not Tested	Percent Exempt
All Students	13.9	19.1	19.4	39.4	8.2	3.6
Females	12.4	19.3	20.4	39.6	5.2	3.1
Males	15.0	18.6	18.0	38.1	6.2	4.1

Table 8-22: 1999 Grade 10 Mathematics Test: Percent Meeting Standards by Ethnic Group

		Meets Standard		Does Not Meet Standard			
Group	Number of Students	Percent Level 4	Percent Level 3	Percent Level 2	Percent Level 1	Percent Not Tested	Percent Exempt
African American/Black	2938	2.0	6.8	11.7	61.8	10.3	7.4
Alaska Native/Native American	1557	3.5	9.9	15.1	55.0	10.4	6.1
Asian/Pacific Islander	5155	17.0	19.0	19.2	36.1	5.2	3.5
Latino/Hispanic	5025	3.3	7.4	11.6	62.2	8.9	6.5
White/Caucasian	53449	15.6	21.2	20.4	34.6	5.0	3.1
Multi-Racial	2665	10.3	16.1	20.3	48.8	4.1	0.4

Table 8-23: 1999 Grade 10 Listening Test: Percent Meeting Standards by Categorical Program

Categorical Program	Number of Students	Percent Meeting Standard	Percent Not Meeting Standard	Percent Not Tested	Percent Exempt
LAP Reading	94	44.7	35.1	18.1	2.1
LAP Mathematics	54	33.3	38.9	22.2	5.6
Title 1 Reading	705	44.0	37.7	15.9	2.4
Title 1 Mathematics	598	43.0	39.8	15.7	1.5
Section 504	396	62.4	23.0	11.4	3.3
Special Education	5205	32.9	44.2	13.1	9.8
Title 1 Migrant Education	411	28.5	54.5	10.9	6.1
Bilingual/ESL	2153	20.8	54.5	13.0	11.8
Gifted/Highly Capable Students	1099	92.9	4.2	2.5	0.5

Table 8-24: Grade 10 Reading Test: Percent Meeting Standards by Categorical Program

		Meets Standard		Does Not Meet Standard			
Categorical Program	Number of Students	Percent Level 4	Percent Level 3	Percent Level 2	Percent Level 1	Percent Not Tested	Percent Exempt
LAP Reading	94	6.4	8.5	31.9	34.0	17.0	2.1
LAP Mathematics	54	3.7	11.1	16.7	42.6	20.4	5.6
Title 1 Reading	705	9.6	10.6	28.1	33.0	16.2	2.4
Title 1 Mathematics	598	10.7	11.0	25.3	33.8	17.7	1.5
Section 504	396	19.4	14.9	27.0	22.5	12.9	3.3
Special Education	5205	4.2	5.9	19.5	44.1	16.5	9.8
Title 1 Migrant Education	411	4.6	9.5	23.1	43.3	12.7	6.8
Bilingual/ESL	2153	1.7	4.2	17.6	49.0	14.6	12.9
Gifted/Highly Capable	1099	71.1	13.8	7.2	2.3	5.2	0.5

Table 8-25: 1999 Grade 10 Writing Test: Percent Meeting Standards by Categorical Program

Categorical Program	Number of Students	Percent Meeting Standard	Percent Not Meeting Standard	Percent Not Tested	Percent Exempt
LAP Reading	94	16.0	55.3	26.6	2.1
LAP Mathematics	54	9.3	51.9	35.2	3.7
Title 1 Reading	705	18.6	55.7	23.1	2.6
Title 1 Mathematics	598	20.2	53.7	24.6	1.5
Section 504	396	23.0	54.5	18.9	3.5
Special Education	5205	6.0	60.5	23.3	10.2
Title 1 Migrant Education	411	10.2	66.9	15.1	7.8
Bilingual/ESL	2153	6.3	57.8	22.2	13.7
Gifted/Highly Capable Students	1099	74.8	18.1	6.6	3.6

Table 8-26: 1999 Grade 10 Mathematics Test: Percent Meeting Standards by Categorical Program

		Meets Standard		Does Not Meet Standard			
Categorical Program	Number of Students	Percent Level 4	Percent Level 3	Percent Level 2	Percent Level 1	Percent Not Tested	Percent Exempt
LAP Reading	94	1.1	3.2	4.3	74.5	13.8	3.2
LAP Mathematics	54	1.9	3.7	3.7	72.2	13.0	5.6
Title 1 Reading	705	1.1	5.8	9.9	68.8	11.3	3.0
Title 1 Mathematics	598	1.2	6.5	9.4	70.2	10.5	2.2
Section 504	396	7.8	11.6	18.2	50.0	8.8	3.5
Special Education	5205	1.2	3.0	6.0	69.8	9.8	10.2
Title 1 Migrant Education	411	1.2	3.2	5.6	74.7	7.8	7.5
Bilingual/ESL	2153	2.4	4.4	8.1	63.3	8.6	13.2
Gifted/Highly Capable	1099	50.0	27.2	12.1	7.6	2.6	0.4

MEAN ITEM PERFORMANCE AND ITEM-TEST CORRELATIONS

As discussed in Part 2, traditional item statistics were used, along with Rasch difficulties and fit statistics, to evaluate the quality of items. All items in the pool were evaluated together and items that met quality standards were retained in the item pool. Mean item performance for multiple choice items can range from 0 to 1. This is often called the p-value. Mean item performance for short-answer items can range from 0 to 2. Mean item performance for extended response items can range from 0 to 4. For the Writing test, mean scores represent the average scores for each of the scoring rules applied to the written piece. There are two written pieces in the Grade 10 WASL. Students can receive from 0 to 4 points for Content, Organization, and Style and from 0 to 2 points for Writing Mechanics for *each* of the written pieces. The higher the mean item performance, the easier the item. Item-test correlations can range from -1.0 to 1.0; positive correlations indicate that item performance is related to overall test performance. Rasch item difficulties can range from –4.0 to 4.0, with negative numbers representing easier items and positive numbers representing more difficult items. The data provided in Tables 8-27 through 8-30 indicate the number of points possible for the items or writing scores, the item or score means, the item score to test score correlations, and the Rasch item difficulties for each of the items in the Listening, Writing, Reading, and Mathematics tests respectively.

Table 8-27: 1999 Grade 10 Listening Test: Number of Points Possible Per Item, Mean Item Performance, Item-Test Correlation, and Rasch Item Difficulty for Each Item

Item Number in Test Booklet	Number Possible	Item Mean	Item-Test Correlation	Rasch Item Difficulty
1	1	.94	.23	-1.45
2	1	.51	.05	1.67
3	1	.94	.18	-1.31
4	1	.85	.32	-0.32
5	1	.76	.19	0.38
6	1	.90	.32	-0.81
7	2	1.41	.21	0.42
8	2	1.15	.22	1.41

Table 8-28: 1999 Grade 10 Writing Test: Number of Points Possible Per Score-Type, Mean Score, and Score-Total Test Correlation for Each Score

Prompt Number	Score Type	Score Points Possible	Score Mean	Score-Total Test Correlation
1	Content, Organization & Style	4	2.56	.59
	Writing Mechanics	2	1.31	.61
2	Content, Organization & Style	4	2.61	.60
	Writing Mechanics	2	1.33	.65

Table 8-29: 1999 Grade 10 Reading Test: Number of Points Possible Per Item, Mean Item Performance, Item-Test Correlation, and Rasch Item Difficulty for Each Item

Item Number in Test Booklet	Points Possible	Item Mean	Item-Test Correlation	Rasch Item Difficulty
1	1	.78	.46	-0.59
2	2	1.48	.38	-0.53
3	1	.86	.43	-1.26
4	1	.79	.28	-0.67
5	1	.58	.28	0.48
6	1	.82	.42	-0.94
7	1	.86	.52	-1.29
8	2	1.47	.45	-0.43
9	1	.88	.37	-1.45
10	1	.70	.36	-0.11
11	1	.55	.36	0.64
12	1	.56	.39	0.58
13	2	1.23	.28	0.58
14	1	.78	.37	-0.61
15	2	1.35	.50	0.23
16	1	.63	.36	0.25
17	1	.68	.36	-0.02
18	1	.67	.47	0.02
19	2	1.27	.58	0.47
20	1	.72	.45	-0.26
21	1	.52	.44	0.78
22	2	1.04	.51	1.01
23	1	.33	.24	1.74
24	4	2.24	.61	0.66
25	1	.82	.49	-0.93
26	1	.88	.48	-1.42
27	1	.84	.43	-1.11
28	2	1.26	.42	0.49
29	1	.54	.47	0.71
30	4	1.46	.51	1.46
31	1	.52	.42	0.79
32	2	1.22	.52	0.63
33	1	.53	.28	0.73
34	1	.64	.45	0.20
35	2	1.36	.48	0.31
36	1	.79	.44	-0.71
37	1	.67	.41	0.06
38	1	.80	.46	-0.77
39	1	.46	.14	1.05
40	2	1.71	.39	-0.75

Table 8-30: 1999 Grade 10 Mathematics Test: Number of Points Possible Per Item, Mean Item Performance, Item-Test Correlation, and Rasch Item Difficulty for Each Item

Item Number in Test Booklet	Points Possible	Item Mean	Item-Test Correlation	Rasch Item Difficulty
1	1	.56	.42	-0.47
2	1	.62	.42	-0.77
3	2	1.19	.56	-0.46
4	1	.27	.33	1.05
5	2	1.46	.52	-1.14
6	1	.57	.48	-0.49
7	4	1.62	.71	0.36
8	1	.29	.46	0.92
9	1	.43	.26	0.20
10	1	.13	.18	2.13
11	2	1.04	.49	-0.21
12	1	.63	.43	-0.80
13	1	.51	.51	-0.23
14	2	1.12	.44	-0.30
15	1	.80	.45	-1.78
16	4	1.75	.66	0.16
17	1	.46	.41	0.01
18	2	.99	.55	0.00
19	1	.20	.22	1.50
20	1	.56	.51	-0.47
21	2	.81	.59	0.38
22	1	.69	.49	-1.14
23	1	.73	.41	-1.34
24	1	.73	.38	-1.37
25	1	.46	.42	0.01
26	2	1.05	.49	-0.10
27	1	.52	.29	-0.25
28	2	.84	.43	0.21
29	4	1.70	.70	0.21
30	1	.46	.38	0.01
31	1	.41	.31	0.27
32	1	.37	.27	0.46
33	2	1.27	.46	-0.66
34	1	.36	.20	0.55
35	1	.29	.42	0.93
36	4	2.31	.62	-0.40
37	2	.65	.61	0.76
38	1	.43	.27	0.20
39	1	.41	.38	0.30
40	2	.73	.62	0.52
41	1	.35	.23	0.61
42	2	1.28	.55	-0.61
43	1	.46	.43	0.02
44	1	.29	.47	0.93
45	1	.40	.46	0.33
46	1	.48	.40	-0.07

APPENDIX A

Washington Essential Academic Learning Requirements in Reading, Writing, Communication, and Mathematics

Reading

1. The student understands and uses different skills and strategies to read.

1.1 Uses word recognition and word meaning skills to read and comprehend text (e.g., phonics, context clues, picture clues, and word origins; roots, prefixes, and suffixes of words)

1.2 Builds vocabulary through reading

1.3 Reads fluently, adjusting reading for purpose and material

1.4 Understands elements of literary (fiction)

1.5 Understands features of non-fiction text and computer software (e.g., titles, headings, pictures, maps, and charts to find and understand specific information)

2. The student understands the meaning of what is read.

2.1 Comprehends important ideas and details

2.2 Expands comprehension by analyzing, synthesizing, and interpreting information and ideas

2.3 Thinks critically about text and analyzes author's use of language, style, purpose, and perspective

3. The student reads different materials for a variety of purposes.

3.1 Reads to learn new information

3.2 Reads to perform tasks

3.3 Reads for literary experience

3.4 Reads for career applications

4. The student sets goals and evaluates progress to improve reading.

4.1 Assesses strengths and need for improvement

4.2 Seeks and offers feedback to improve reading

4.3 Develops interests and shares reading experiences

Writing

1. The student writes clearly and effectively

1.1 Develops concept and design (develops a topic or theme; organizes written thoughts with a clear beginning, middle, and end; uses transitional sentences and phrases to connect ideas; writes coherently and effectively)

1.2 Uses style appropriate to audience and purpose (uses voice, word choice, and sentence fluency for intended style and audience)

1.3 Applies writing conventions (grammar, punctuation, capitalization)

2. The student writes in a variety of forms for different audiences and purposes.

2.1 Writes for different audiences

2.2 Writes for different purposes (telling stories, presenting analytical responses to literature, persuading, conveying technical information, completing a team project, explaining concepts and procedures)

2.3 Writes in a variety of forms (narratives, journals, poems, essays, stories, research reports, and technical writing)

3. The student understands and uses the steps of the writing process.

3.1 Prewrites (generates ideas and gather information for writing)

3.2 Drafts (elaborates on a topic and supporting ideas)

3.3 Revises (collects input and enhances style and text)

3.4 Edits (uses resources to correct spelling, punctuation, grammar, and usage)

3.5 Publishes (selects publishing form and produces a completed writing project to share with a chosen audience)

4. The student analyzes and evaluates the effectiveness of written work.

4.1 Assesses own strengths and needs for improvement (analyzes effectiveness of own writing and sets goals for improvement)

4.2 Seeks and offers feedback

Communication

1. **The student uses listening and observing skills to gain understanding.**
   - 1.1 Focuses attention
   - 1.2 Listens and observes to gain and interpret information
   - 1.3 Checks for understanding by asking questions and paraphrasing

2. **The student communicates ideas clearly and effectively.**
   - 2.1 Communicates clearly to a range of audiences for different purposes
   - 2.2 Develops content and ideas (develops a topic or theme; organizes thoughts around a clear beginning, middle, and end; uses transitional sentences and phrases to connect ideas; speaks coherently and effectively)
   - 2.3 Uses effective delivery (adjusts speaking strategies for a variety of audiences and purposes by varying tone, pitch, projection, posture, eye contact, facial expressions body language, and pace of speech to create effect and aid communication)
   - 2.4 Uses effective language and style (uses language that is grammatically correct, precise, engaging, and well suited to topic, audience and purpose)
   - 2.5 Effectively uses action, sound, and/or images to support presentations

3. **The student uses communication strategies and skills to work effectively with others.**
   - 3.1 Uses language to interact effectively and responsibly with others
   - 3.2 Works cooperatively as a member of a group
   - 3.3 Seeks agreement and solutions through discussion

4. **The student analyzes and evaluates the effectiveness of formal and informal communication.**
   - 4.1 Assess strengths and needs for improvement (analyzes effectiveness of own writing and sets goals for improvement)
   - 4.2 Seeks and offers feedback (seeks and uses feedback to improve communication; offers suggestions and comments to others)
   - 4.3 Analyzes mass communication
   - 4.4 Analyzes how communication is used in career settings

Mathematics

1. The student understands and applies the concepts and procedures of mathematics.

1.1 Understands and applies concepts and procedures of number sense (number and numeration, number theory, computation, and estimation)

1.2 Understands and applies concepts and procedures of measurement (attributes and dimensions, approximation and precision, systems and tools)

1.3 Understands and applies concepts and procedures of geometric sense (shape and dimension, relationships, and transformation)

1.4 Understands and applies concepts and procedures of probability and statistics (probability, statistics, prediction, and inference)

1.5 Understands and applies concepts and procedures of algebraic sense (patterns, relations, representations, and operations)

2. The student uses mathematics to define and solve problems.

2.1 Investigates situations (by searching for patterns and exploring a variety of approaches)

2.2 Formulates questions and defines problems

2.3 Constructs solutions (by choosing necessary information and using the appropriate tools, concepts and procedures)

3. The student uses mathematical reasoning.

3.1 Analyzes information (from a variety of sources; uses models, known facts, patterns, and relationships to validate thinking)

3.2 Predicts results and makes inferences and conjectures based on analysis of problem situations

3.3 Draws conclusions and verifies results (supports mathematical arguments, justifies results, and checks for reasonableness of solutions)

4. The student uses communicates knowledge and understanding in both everyday and mathematical language.

4.1 Gathers information (reads, listens, and observes to extract mathematical information)

4.2 Organizes and interprets information

4.3 Represents and shares mathematical information (shares, explains, defends mathematical ideas using terms, language, charts, and graphs, etc. that can be clearly understood by a variety of audiences

Mathematics (Cont.)

4. The student understands how mathematical ideas connect within mathematics, to other subject areas, and to real life situations.

5.1 Relates ideas and concepts within mathematics (recognizes relationships among mathematical ideas and topics)

5.2 Relates mathematical concepts and procedures to other disciplines (identifies and applies mathematical thinking and notation in other subject areas

5.3 Relates mathematical concepts and procedures to real-life situations (understands the connections between mathematics and problem solving skills used every day at work and at home)

APPENDIX B

WASHINGTON ASSESSMENT OF STUDENT LEARNING

GRADE 10

MATHEMATICS TEST SPECIFICATIONS

Test Specifications for the
Washington Assessment of Student Learning
Grade 10 Mathematics
February, 1999

I. Test Purpose

The purpose of this test is to measure the level of mathematics proficiency that Washington students have achieved by the spring of the tenth grade, according to the Essential Academic Learning Requirements established by the Washington Commission on Student Learning (CSL). The Essential Academic Learning Requirements (EALRs) consist of the mathematics concepts and procedures and four fundamental processes (solving problems, reasoning, communicating, and making connections). These concepts and procedures and the processes are grouped into the following content and process strands:

Content Strands	Process Strands
NS=Number Sense	SP=Solving Problems
ME=Measurement	RL=Reasoning Logically
GS=Geometric Sense	CU=Communicating Understanding
PS=Probability and Statistics	MC=Making Connections
AS=Algebraic Sense	

In keeping with the CSL Essential Academic Learning Requirements Technical Manual (Washington State Commission on Student Learning, July 17, 1998), these Essential Academic Learning Requirements in mathematics (the content and the process strands) are viewed as an integrated whole. Each test item will be identified as to the primary content and/or process strand it is assessing.

The following learning targets are intended to summarize the knowledge or Essential Academic Learning Requirements as identified in the mathematics section of the Essential Academic Learning Requirements Technical Manual (Washington State Commission on Student Learning, July 17, 1998). These benchmarks are identified by numbers in parentheses after each target and are listed at the end of this document.

II. Content Strands and Learning Targets*

<u>Strand 1: Number Sense</u>

NS01 (Numbers and Numeration)

Demonstrates an understanding of and uses the symbolic representations of real numbers; explains the magnitude of numbers by comparing and ordering real numbers. (1.1.1, 1.1.2)**

NS02 (Number Theory)

Demonstrates an understanding of concepts and processes involving primes, composite numbers, divisibility, and factors and multiples; understands the properties of the system of real numbers. (1.1.3, 1.1.1)

*** The first digit in this number corresponds to the Essential Academic Learning Requirement; the second digit corresponds to its component; the third digit corresponds to its specific benchmark, as presented in the Essential Academic Learning Requirements Technical Manual. To locate the specific wording for any item listed here, refer to the content area section in the Technical Manual.*

NS03 (Conceptual Understanding of Operations)

Demonstrates an understanding of operations on real numbers. (1.1.5)

NS04 (Estimation)

Identifies situations in which estimation is sufficient and computation is not required; uses estimation to predict computation results and to determine reasonableness of answers. (1.1.8, 1.1.9)

NS05 (Computation)

Computes with real numbers, powers, and roots. (1.1.6)

NS06 (Ratio and Proportion)

Demonstrates an understanding and applies concepts of ratio and both direct and inverse variation. (1.1.4)

CBE (Classroom-Based Evidence)

Use mental arithmetic, pencil and paper, calculator or computer as appropriate to the task. (1.1.7)

Strand 2: Measurement

ME01 (Attributes and Dimensions)

Demonstrates an understanding of how changes in dimension affect perimeter, area, and volume; calculates rate and other derived and indirect measurements. (1.2.1, 1.2.3)

ME02 (Approximation/Precision)

Demonstrates an understanding that the precision and accuracy of measurement is affected by the measurement tools and calculating procedures; uses estimation to obtain reasonable approximations. (1.2.4, 1.2.5)

ME03 (Calculation)

Measures objects and events directly or uses indirect methods; calculates rate and other derived and indirect measurements. (1.2.2, 1.2.3)

ME04 (Systems)

Demonstrates an understanding of the benefits of standard units of measurement and the advantages of the metric system; compares, contrasts, and uses both the U.S. and metric systems; selects and uses tools that will provide an appropriate degree of precision. (1.2.6, 1.2.7, 1.2.8)

Strand 3: Geometric Sense

GS01 (Shapes and Figures)

Compares, describes, and classifies 2- and 3-dimensional geometric figures; constructs geometric models and scale drawings using tools as appropriate. (1.3.1, 1.3.2)

GS02 (Locations)

Demonstrates an understanding of and uses coordinate grids. (1.3.3)

GS03 (Geometric Properties)

Demonstrates an understanding of simple differences between the geometric properties of a plane and a sphere and uses properties of symmetry, similarity, and congruence. (1.3.4, 1.3.5)

GS04 (Transformations)

Demonstrates an understanding of and constructs multiple geometric transformations. (1.3.6)

CBE (Classroom-Based Evidence)

Constructs geometric models and scale drawings using tools when appropriate. (1.3.2)

Strand 4: Probability and Statistics

PS01 (Determine Probabilities)

Demonstrates an understanding of the properties of dependent and independent events; understands and uses appropriate counting procedures to determine probabilities; uses both experimental and theoretical methods to determine probabilities. (1.4.1, 1.4.2, 1.4.3)

PS02 (Collect and Organize Data)

Demonstrates an understanding of how to collect data using appropriate methods and technology; organizes and displays data in appropriate forms such as tables, graphs, scatter plots, and box plots. (1.4.5, 1.4.6)

PS03 (Describe and Analyze Data)

Calculates and uses the different measures of central tendency, variability, and range as appropriate in describing a set of data; understands that there can be different interpretations of the same set of data and uses statistics to support different points of view. (1.4.4, 1.4.7)

PS04 (Make Inferences and Predictions)

Designs and conducts experiments to verify or disprove predictions; understands and makes inferences based on analysis or experimental results. (1.4.8, 1.4.9)

CBE (Classroom-Based Evidence)

Demonstrates an understanding of and can use appropriate methods and technology for collecting data. (1.4.5)

Strand 5: Algebraic Sense

AS01 (Patterns, Series, and Sequences)

Recognizes, creates, extends, and generalizes patterns, series, and sequences. (1.5.1)

AS02 (Represents Functions and Relationships)

Understands, develops, and expresses rules for patterns; translates among tabular, symbolic, and graphical representations of relations. (1.5.2, 1.5.3)

AS03 (Equations)

Creates and solves equations and inequalities; evaluates and simplifies expressions. (1.5.5, 1.5.6)

AS04 (Application of Equations)

Represents situations that involve variable quantities with expressions, formulas and equations, and inequalities. (1.5.4)

III. Process Strands and Learning Targets

Strand 6: Solving Problems

SP01 (Investigates Situations)

Analyzes and uses multiple strategies; identifies missing/extraneous information and compensates for it; and searches systematically for patterns in complex situations. (2.1.1, 2.1.2, 2.1.3)

SP02 (Defines the Problem)

Defines the problem to be solved in complex situations; identifies unknowns and questions to be answered in complex situations. (2.2.1, 2.2.2, 2.2.3)

SP03 (Constructs Solutions)

Organizes and synthesizes information from multiple sources; selects and uses appropriate tools, methods, operations, and processes to construct solutions. (2.3.1, 2.3.2, 2.3.3)

CBE (Classroom-Based Evidence)

- Recognizes when an attempted approach is unproductive and tries to modify it or tries a new approach. (2.1.4)
- Organizes relevant information collected from a variety of sources. (2.3.1)

Strand 7: Reasoning Logically

RL01 (Analyzes Information)

Interprets and integrates information from multiple sources. (3.1.1)

RL02 (Predicts)

Makes and explains conjectures and inferences based on analysis of problem situations. (3.2.1)

RL03 (Verifies)

Validates thinking and mathematical ideas using models, known facts, patterns, relationships, counter-examples, and proportional reasoning; tests conjectures and inferences by formulating a proof or by constructing a counter-example; supports arguments and justifies results using inductive and deductive reasoning; checks for reasonableness of results; evaluates and reflects on procedures and results and makes necessary revisions. (3.1.2, 3.3.1, 3.3.2, 3.3.3, 3.3.4)

Strand 8: Communicating Understanding

CU01 (Gathering Information)

Develops or selects an efficient system for collecting information; uses reading, listening, and observation skills to access and extract mathematical information from multiple sources. (4.1.1, 4.1.2)

CU02 (Organizing and Interpreting & Representing and Sharing)

Organizes, clarifies, and refines mathematical information in multiple ways; expresses complex ideas and situations using mathematical language and notation in appropriate and efficient forms; clearly and effectively expresses/presents mathematical ideas using both everyday and mathematical language which is appropriate to the audience. (4.2.1, 4.3.1, 4.3.2)

CBE (Classroom-Based Evidence)

- Uses reading, listening, and observation skills to access and extract mathematical information from multiple sources such as pictures, diagrams, physical models, oral narratives, and symbolic representations. (4.1.2)
- Chooses appropriate technology to browse, select, and retrieve mathematical

Strand 9: Making Connections*

MC01 (Connections among Concepts and Procedures and between Different Mathematical Representations)

Relates and uses conceptual and procedural understandings among multiple mathematical content areas or relates and uses multiple equivalent mathematical models and representations. (5.1.1, 5.1.2)

CBE (Classroom-Based Evidence)

Extends mathematical patterns and ideas to other disciplines. (5.2.1)

IV. Content Organization

The tenth-grade mathematics test will consist of 46 items or questions, resulting in 70 points per form. Items will be written at a reading level appropriate to a tenth-grade audience; thus, item development will aim for eighth-grade readability. Test forms will include the following item types:

- Multiple-choice items: The student will have four responses from which to choose: the correct answer and three distractors. There will be 30 multiple-choice items per form, each worth 1 point.
- Short-answer items (including enhanced multiple-choice*): The student must construct a short response. For example, the student may be required to write a sentence or equation; complete a table, graph, or chart; draw a picture; construct a diagram; or carry out a calculation. There will be 12 short-answer items per form, each worth 2 points. No more than 25% of the short-answer items in a test form will be enhanced multiple-choice items.
- Extended-response items: The student must construct a longer response than that for a short-answer item. For example, the student may be required to create a graph showing the appropriate data, title, and labeled axes; create and/or extend tables, diagrams, or pictures; provide a lengthy written explanation, a written explanation with equations, pictures, and/or diagrams. There will be 4 extended-response items per form, each worth 4 points.

Multiple-choice and short-answer items will be used to assess targets in the content strands. All three-item types will be used to assess targets in the process strands. Each item in a process strand will indicate the appropriate content target, if any, in its item code.

The test will be administered in two parts. It is intended that each of the two parts of the test will contain about 23 items in approximately the following proportions: 15 multiple choice, 6 short answer, and 2 extended response.

The test will be administered in two separate sessions, each of which will be about 1 hour and 15 minutes long, plus a 15-minute break time that students may take. Though the test is not specifically a timed test, total testing time for standard administration should take about 2 1/2 hours.

Each test form will contain a variety of items so that all strands or Essential Academic Learning Requirements are addressed. The two parts of the test will be constructed so as to separate the items for which tools (such as rulers or calculators) must not be used from the items for which tools are encouraged.

Types of tools that may be used for the "tools section" of the assessment include the following: calculators, protractors, rulers, pattern blocks, and other classroom manipulatives.

V. Test Scoring

Each multiple-choice item is worth 1 point, each short-answer item is worth 2 points, and each extended-response item is worth 4 points. Thus, for example, in a 46-item test, 30 multiple-choice items would be worth 30 points, 12 short-answer items would be worth 24 points, and 4 extended-response items would be worth 16 points, making a total of 70 possible points. Multiple-choice items would account for 43% of the total score, while the constructed-response items (both short-answer and extended-response items) would account for 57% of the total points. This distribution is shown below:

Type	Number of Items	Total Points	Percent of the Total Score
Multiple-choice	30	30	43%
Short-answer*	12	24	34%
Extended-response	4	16	23%
Total	46	70	100%

**Note: No more than 3 short-answer items will be enhanced multiple-choice items.*

Scoring of Open-Ended Items

Individual scoring criteria will be developed for each constructed-response item. Short-answer items will be scored on a scale of 0 to 2 points, and extended-response items will be scored on a scale of 0 to 4 points. The following scoring criteria are generalized to include either content or process aspects.

General scoring criteria for short-answer mathematical concepts and procedures items

2 points—Student's response shows complete understanding of the concept or task, as well as consistent and correct use of applicable information and/or procedures. Setup and computations are accurate.

1 point—Student's response shows partial understanding of the concept or task. There may be minor errors in the use of applicable information and/or procedures. Setup or computations may have minor errors.

0 points—Student's response shows very little or no understanding of the concept or task; or the prompt may simply be recopied, or the response may indicate "I don't know" or a question mark (?).

General scoring criteria for short-answer solving problems items

2 points—Student's response shows thorough investigation, clear understanding of the problem, and/or effective and viable solution.

1 point—Student's response shows partial investigation and/or understanding of the problem, and/or a partially complete or partially accurate solution.

0 points—Student's response shows very little or no investigation and/or understanding of the problem, and/or no visible solution; or the prompt may simply be recopied, or may indicate "I don't know" or a question mark (?).

General scoring criteria for short-answer mathematical reasoning items

2 points—Student's response shows effective reasoning through a complete analysis or thorough interpretation, supported predictions, and/or verification.

1 point—Student's response shows somewhat flawed reasoning either through incomplete analysis or interpretation, prediction that lacks support, or inadequate/incomplete verification.

0 points—Student's response shows very little or no evidence of reasoning; or the prompt may simply be recopied, or may indicate "I don't know" or a question mark (?).

General scoring criteria for short-answer mathematical communication items

2 points—Student's response shows understanding of how to effectively and appropriately interpret, organize, and/or represent mathematical information relevant to the concept.

1 point—Student's response shows some understanding of how to interpret, organize, and/or represent mathematical information relevant to the concept; however, the response is not complete or effectively presented.

0 points—Student's response shows very little or no understanding of how to interpret, organize and/or represent mathematical information relevant to the concept; or the prompt may simply be recopied, or the response may indicate "I don't know" or a question mark (?).

General scoring criteria for short-answer making mathematical connections items

2 points—Student's response makes clear and effective connections within and/or between conceptual or procedural areas or between different mathematical representations.

1 point—Student's response makes vague or partially accurate connections within and/or between conceptual or procedural areas or between different mathematical representations.

0 points—Student's response makes little or no connection within or between conceptual or procedural areas or between different mathematical representations; or the prompt may simply be recopied, or may indicate "I don't know" or a question mark (?).

General scoring criteria for extended-response solving problems items

4 points—Student's response shows clear and accurate understanding of the task and how to effectively search for patterns, identify missing or extraneous information, formulate questions and define the problem, and/or select appropriate tools and strategies to develop an effective solution, organizing relevant information and use appropriate mathematical concepts and procedures.

3 points—Student's response shows understanding of the task and how to search for patterns, identify most of the missing or extraneous information, formulate some questions and begin to define the problem, and/or select tools and strategies to develop a solution, organizing information and using correct mathematical concepts and procedures. Minor errors or gaps limit viability of solutions.

2 points—Student's response shows some understanding of the task and how to search for patterns, identify some missing or extraneous information, formulate a question, and/or select a tool and/or a strategy to begin developing a solution, using mostly appropriate mathematical concepts and procedures. Gaps or errors limit viability of solutions.

1 point—Student's response shows little understanding of the task or how to search for patterns, identify missing or extraneous information, formulate questions, and/or select appropriate tools and strategies to develop a solution. Major gaps, errors, or poor conceptual understanding prevent a viable solution.

0 points—Student's response provides no evidence of problem-solving skills or shows very little or no understanding of the task; or the prompt may simply be recopied, or the response may indicate "I don't know" or a question mark (?).

General scoring criteria for extended-response mathematical reasoning items

4 points—Student's response shows effective interpretations, comparisons, or contrasts of information from sources; effective use of examples, models, facts, patterns, or relationships to validate and support reasoning; insightful conjectures and inferences; and/or systematic and successful evaluation of effectiveness of procedures and results; with effective support for arguments and results.

3 points—Student's response shows partially effective interpretations, comparisons, or contrasts of information from sources; use of examples, models, facts, patterns, or relationships to validate and support reasoning; expected conjectures and inferences; and/or mostly successful evaluation of effectiveness of procedures and results with acceptable support for arguments and results.

2 points—Student's response shows routine interpretations, comparisons, or contrasts of information from sources; examples, models, facts, patterns, or relationships which partially validate and support reasoning; naive conjectures and inferences; and/or partial evaluation effectiveness of procedures and results with partial support for arguments and results.

1 point—Student's response shows an attempt to interpret, compare, or contrast information from sources; examples, models, facts, patterns, or relationships may not be included to validate or support reasoning; naive conjectures and inferences; and/or attention to wrong information or persistence with faulty strategy when evaluating effectiveness of procedures and results.

0 points— Student's response shows very little or no evidence of reasoning; or the prompt may simply be recopied, or may indicate 'I don't know' or a question mark (?).

General scoring criteria for extended-response mathematical communication items

4 points—Student gathers all applicable information from appropriate sources; demonstrates interpretations and understandings in a clear, systematic, and organized manner; represents mathematical information and ideas in an effective format for the task, situation, and audience.

3 points—Student gathers applicable information from appropriate sources; demonstrates interpretations and understandings in a clear and organized manner; represents mathematical information and ideas in an expected format for the task, situation, and audience.

2 points—Student gathers information from appropriate sources; demonstrates interpretations and understandings in an understandable manner; represents mathematical information and ideas in an acceptable format for the task, situation, and audience.

1 point—Student gathers little information from appropriate sources; demonstrates interpretations and understandings in a manner that may be disorganized or difficult to understand; represents mathematical information and ideas in a format that may be inappropriate for the task, situation, and audience.

0 points—Student's response shows very little or no understanding of how to interpret, organize and/or represent mathematical information relevant to the concept; or the prompt may simply be recopied, or may indicate "I don't know" or a question mark (?).

VI. Reporting Schema and Item Distribution

All Essential Academic Learning Requirements (the content and process strands) will be addressed in each test form. Items will be classified according to their primary content or process strands. The overall item distribution in an operational test form is intended to look as follows:

Essential Academic Learning Requirement	Strand	Number of Targets	Number of Items (approximate range)	Number of Points (approximate range)
Concepts and Procedures	Number Sense	6	4-8	5-10
	Measurement	4	4-8	5-10
	Geometric Sense	4	4-8	5-10
	Probability and Statistics	4	4-8	5-10
	Algebraic Sense	4	4-8	5-10
Content total				Approx. 36 pts
Solving Problems		3	2-6	6-14
Reasoning Logically		3	2-5	5-12
Communicating Understanding		2	2-5	5-10
Making Connections		1	2-4	3-6
Process total				Approx. 34 pts

Another distribution according to item type is shown below:

Essential Academic Learning Requirement	Strand	MC Items	SA Items	ER Items	Total No. of Items (range)
			Items	Items	
Concepts and Procedures	Number Sense (6)*	3-6	1-2	0	4-8
	Measurement (4)	3-6	1-2	0	4-8
	Geometric Sense (4)	3-6	1-2	0	4-8
	Probability and	3-6	1-2	0	4-8
	Statistics (4)				
	Algebraic Sense (4)	3-6	1-2	0	4-8
Solving Problems (3)		0-2	1-2	1-2	2-6
Reasoning Logically (3)		0-2	1-2	1-2	2-5
Communicating		0-2	1-2	1-2	2-5
Understanding (2)					
Making Connections (1)		1-2	1-2	0	2-4
Total Number of Items		30	12	4	
Total Number of Points		30	24	16	

*Numbers in parentheses represent the numbers of learning targets in each strand

VII. General Considerations

Each multiple-choice item will have four responses: the correct answer and three distractors. Distractors will be developed based on the types of errors students are most likely to make.

For test development purposes, item codes will accompany each item and will provide information regarding the content and/or process strand addressed, learning target addressed, item format, and correct answer key. The following abbreviations will be used to indicate content and process strands in the codes:

NS=Number Sense
ME=Measurement
GS=Geometric Sense
PS=Probability and Statistics
AS=Algebraic Sense
SP=Solving Problems
RL=Reasoning Logically
CU=Communicating Understanding
MC=Making Connections

The scoring criteria will focus on the clear communication of mathematical ideas, information, and solutions, and will disregard conventions of writing (sentence structure, word choice, usage, grammar, spelling, and mechanics), as long as the wording of the response does not interfere with the mathematical communication.

All items will avoid language that shows bias, offends, or disadvantages a particular group of students. That is, items will not display unfair representations of genders, races, persons with disabilities, or cultural and religious groups.

VIII. Notational Considerations (FOR STEMS)

- In the item stems, numbers (other than years) having more than three digits to the left of the decimal point will include commas to group digits in the usual manner (e.g., 435,000).
- Physical quantities such as length should generally be labeled or accompanied by units. Standard abbreviations (e.g., cm or cu ft) may be used. The unit should be spelled out if any confusion is reasonably possible.
- Variables will always be italicized.
- Parentheses and the symbol · can be used to indicate multiplication.
- The symbols ÷ and a fraction bar can be used to indicate division.
- Fractions will have horizontal lines separating numerator and denominator.
- Integer exponents (no fractional exponents) can be used.
- Large numbers may be used in certain contexts, such as in a chart or graph with a heading "in millions."

IX. Characteristics of Items and Item Stems or Foils

- Each item begins with a stem that asks a question. A stem will usually ask a direct question. It will seldom use an incomplete sentence, be worded negatively, or ask for a "best" answer.
- A stem that gives information might precede a question or a set of questions. A stem may consist of brief written material and/or a graphic, such as a simple diagram, graph, chart, table, or drawing.
- Stems for items will be factually correct and have a readability targeted for an eighth-grade readability. Stems should be adapted specifically for the test. A test item will focus on what is essential and consequential in the stem and will minimize the impact of, or need for,

outside knowledge. The amount of reading will be kept to a minimum so that each item is clear and precise.

- Character names on each form will be representative of the ethnic diversity of Washington students. The names will generally be short and simple to read.
- To the extent possible, no stem, or response for one item will serve as a clue to the correct response for another item.
- Graphs, tables, or figures must be clearly associated with their intended items. Graphics will appear either on the same page as the stimulus or on the facing page. If there is any reasonable chance of confusion, page references will direct students to look at the appropriate graphic.
- Test items will be independent in the sense that the answer for any test item does not depend on knowing the correct answer to another item, so items are not "linked." Note: Linkage will be avoided among different items, not necessarily to parts within a single item. For instance, an enhanced multiple-choice may ask students to explain their reason for selecting a particular response. This is not linking between items.
- When appropriate, several items may center around a particular stem, graph, chart, or scenario, in which case, these items will generally appear on the same page or facing page from the stem.
- All items must clearly indicate what is expected in a response and must help students focus their response. General directions that allow the student more freedom in response format may read as follows: "Show or explain your thinking using words, numbers, and/or diagrams."
- Items in each form are to be balanced by gender and should be gender-neutral for active/passive roles.
- Items testing application and problem solving will involve understandable, realistic situations.
- Pictorial representations will be realistic and authentic for tenth graders.
- On items for which manipulatives and/or tools are encouraged or required, students may be given the opportunity to use any punch-out or overlay manipulatives provided, or may use those classroom manipulatives or tools with which they are most familiar/comfortable, as long as nothing about the tools would introduce bias into the results.
- The tools may include calculators, protractors, rulers, pattern blocks, and any other classroom manipulatives.
- All multiple-choice responses (key and distractors) are to be similar in length and in syntax; students should not be able to rule out a wrong answer or identify a correct response simply by virtue of its looking or sounding different.
- Distractors are to be created so that students must think their way to the correct answer rather than simply identify incorrect responses by virtue of a distractor’s obviously inappropriate nature.
- Distractors should always be plausible (but of course incorrect) in the context of the item stem.
- The responses or distractors will be arranged in a logical order, i.e., numerical or chronological order or according to the length of the distractors.
- Answer keys will be approximately equally distributed among As, Bs, Cs and Ds.

- The responses "Both of the above," "All of the above," "None of the above," and "Neither of the above" will not be used, and the use of the word not will generally be avoided in the item stem.
- Care will be taken not to use items for which wrong methods yield the correct response. For example, "Simplify the fraction 64/16" is a poor item, since the correct response can be obtained by canceling the two sixes.
- If a question is stated in terms of one measurement system, all response options should be given in terms of the same measurement system. Units do not have to be included in the stem, but they should appear in every distractor or response when appropriate.

Needs information about general rules for short-answer and extended response.

- *Prompts for short-answer and extended response items must give students all requirements for successful response.*
- No more than three short-answer items may be enhanced multiple-choice items (in which students are required to choose from four responses and then explain their choice).

Essential Academic Learning Requirements Technical Manual (Washington State Commission on Student Learning, July 17, 1998)

1.1 Understand and Apply Concepts and Procedures from Number Sense

Number and Numeration

1.1.1 Understand and use properties and symbolic representations of real numbers.

1.1.2 Explain the magnitude of numbers by comparing and ordering real numbers.

1.1.3 Understand concepts and use of processes involving primes, composite numbers, divisibility, and factors and multiples.

1.1.4 Understand and apply concepts of ratio and both direct and inverse variation.

Computation

1.1.5 Understand operations on real numbers.

1.1.6 Compute with real numbers, powers, and roots.

1.1.7 Use mental arithmetic, pencil and paper, calculator or computer as appropriate to the task involving real numbers.

Estimation

1.1.8 Identifies situations involving real numbers in which estimation is sufficient and computation is not required.

1.1.9 Use estimation to predict computation results and to determine reasonableness of answers involving real numbers, for example, estimating the interest on a loan.

1.2 Understand and Apply Concepts and Procedures from Measurement

Attributes and Dimensions

1.2.1 Understand how changes in dimension affect perimeter, area, and volume.

1.2.2 Measure objects and events directly or use indirect methods, such as finding the volume of a cone given its height and diameter.

1.2.3 Calculate rate and other derived and indirect measurements.

Approximation and Precision

1.2.4 Understand that the precision and accuracy of measurement is affected by the measurement tools and calculating procedures.

1.2.5 Use estimation to obtain reasonable approximations, for example, estimating how much paint is needed to paint the walls of a classroom.

Systems and Tools

1.2.6 Understand the benefits of standard units of measurement and the advantages of the metric system.

1.2.7 Compare, contrast, and use both the U.S. customary and metric systems.

1.2.8 Select and use tools that will provide an appropriate degree of precision, for example, using kilometers vs. light years.

1.3 Understand and Apply Concepts and Procedures from Geometric Sense

Shape and Dimension

1.3.1 Compare, describe, and classify 2- and 3-dimensional geometric figures.

1.3.2 Construct geometric models and scale drawings using tools as appropriate, for example, designing a house plan or building a model of a bridge.

Relationships and Transformations

1.3.3 Understand and use coordinate grids.

1.3.4 Identify simple differences between geometric properties of a plane and a sphere.

1.3.5 Understand and use properties of symmetry, similarity, and congruence.

1.3.6 Understand and construct multiple geometric transformations using combinations of translation, reflection, or rotation.

1.3.7 Use a variety of tools and technologies to perform geometric constructions.

1.4 Understand and Apply Concepts and Procedures from Probability and Statistics

Probability

1.4.1 Understand the properties of dependent and independent events.

1.4.2 Understand and use appropriate counting procedures to determine probabilities.

1.4.3 Use both experimental and theoretical methods to determine probabilities.

Statistics

1.4.4 Use statistics to support different points of view, for example, in a debate or a position paper.

1.4.5 Collect data using appropriate methods and technology.

1.4.6 Organize and display data in appropriate forms, such as tables, graphs, scatter plots, and box plots.

1.4.7 Calculate and use the different measures of central tendency, variability, and range as appropriate in describing sets of data.

Prediction and Inference

1.4.8 Design and conduct experiments to verify or disprove predictions.

1.4.9 Understand and make inferences based on the analysis of experimental results.

1.5 Understand and Apply Procedures from Algebraic Sense

Relations and Representations

1.5.1 Recognize, create, extend, and generalize patterns, series and sequences.

1.5.2 Understand, develop, and express rules describing patterns.

1.5.3 Translate among tabular, symbolic, and graphical representations of relations, for example, displaying information from a table as a graph.

1.5.4 Represent situations that involve variable quantities with expressions, formulas and equations, and inequalities.

Operations

1.5.5 Evaluate and simplify expressions.

1.5.6 Create and solve equations and inequalities.

2.0 Define and Solve Problems

Investigate Situations

2.1.1 Search systematically for patterns in complex situations.

2.1.2 Analyze and use multiple strategies.

2.1.3 Identify what information is missing or extraneous and compensate for it.

2.1.4 Analyze an unproductive approach and attempt to modify it or try a new approach.

Formulate Questions and Define the Problem

2.2.1 Identify questions to be answered in complex situations.

2.2.2 Define problems in complex situations.

2.2.3 Identify the unknowns in complex situations.

Construct Solutions

2.3.1 Organize and analyze information from multiple sources.

2.3.2 Select and use appropriate mathematical tools.

2.3.3 Apply appropriate methods, operations, and procedures to construct a solution.

3.0 Use Mathematical Reasoning

Analyze Situations

3.1.1 Interpret and integrate information from multiple sources.

3.1.2 Validate thinking and mathematical ideas using models, known facts, patterns, relationships, counterexamples, and proportional reasoning.

Predict Results and Make Inferences

3.2.1 Make and explain conjectures and inferences based on analysis of problem situations.

Draw Conclusions and Verify Results

3.3.1 Test conjectures and inferences by formulating a proof or by constructing a counterexample.

3.3.2 Support arguments and justify results using deductive and inductive reasoning.

3.3.3 Check for reasonableness of results.

3.3.4 Reflect on and evaluate procedures and results and make necessary revisions.

4.0 Communicate Knowledge and Understanding

Gather Information

4.1.1 Develop or select an efficient system for collecting information.

4.1.2 Use reading, listening, and observation skills to access and extract mathematical information from multiple, self-selected sources such as pictures, diagrams, hysical models, oral narratives, and symbolic representations.

4.1.3 Integrate the use of a variety of technologies to browse, select, and retrieve mathematical information from multiple sources.

Organize and Interpret Information

4.2.1 Organize, clarify, and refine mathematical information in multiple ways: reflecting, verbalizing, discussing, or writing.

Represent and Share Information

4.3.1 Express complex ideas and situations using mathematical language and notation in appropriate and efficient forms.

4.3.2 Express or present mathematical ideas clearly and effectively using both everyday and mathematical language appropriate to an audience.

5.0 Understand Mathematical Connections

Relate Concepts and Procedures Within Mathematics

5.1.1 Relate and use conceptual and procedural understandings among multiple mathematical content areas.

5.1.2 Relate and use multiple equivalent mathematical models and representations.

Relate Mathematical Concepts and Procedures to Other Disciplines

5.2.1 Extend mathematical patterns and ideas in other disciplines.

5.2.2 Apply mathematical thinking and modeling in other disciplines.

5.2.3 Describe examples of contributions to the development of mathematics such as the contributions of men, women, and different cultures.

Relate Mathematical Concepts and Procedures to Real-life Situations

5.3.1 Identify situations in which mathematics can be used to solve problems with local, national, or international implications, such as calculating resources necessary for interstate highway maintenance.

5.3.2 Investigate the mathematical knowledge and training requirements for occupational/career areas of interest.

Item Specifications for the
Washington Assessment of Student Learning
Grade 10 Mathematics
February, 1999

The goal of this test is to measure the level of mathematics proficiency that Washington students have achieved by the spring of the tenth grade. To thoughtfully and equitably achieve this goal, there are three response formats: multiple choice, short answer, and extended response. The test items will assess proficiency in mathematics according to the Essential Academic Learning Requirements (EALRs) established by the Washington Commission on Student Learning (CSL). For classification and reporting purposes, each test item will be correlated to one of the content or process strands listed below:

Content Strands
1. Number Sense
2. Measurement
3. Geometric Sense
4. Probability & Statistics
5. Algebraic Sense

Process Strands
6. Solving Problems
7. Reasoning Logically
8. Communicating Understanding
9. Making Connections

There are some general considerations that can make the tasks of item and assessment writing more efficient and effective. These considerations include, but are not limited to, the following:

- Each test form will contain items that assess learning targets from all of the content and process strands.
- Test items that assess each learning target will not be limited to one particular type of response format. However, extended-response formats will be reserved for those items that assess learning targets in the process strands.
- Test questions should focus on real-world situations to which a tenth-grade student could relate.
- All items must clearly indicate what is expected in a response and must help students focus their responses. The test items should be worded precisely and clearly. The better focused an item, the more reliable and fair it is certain to be, and the more likely all students will understand what is required of them.
- All items will avoid language that shows bias, offends, or disadvantages a particular group of students. That is, items will not display unfair representations of genders, races, persons with disabilities, or cultural and religious groups.
- All multiple-choice responses (key and distractors) will be similar in length and in syntax. Students should not be able to rule out a wrong answer or identify a correct response simply by virtue of its looking or sounding different. Also, distractors should be created so that students must think their way to the correct answer rather than simply identify incorrect responses by virtue of a distractor's obviously inappropriate nature. Distractors should always be plausible (but of course incorrect) in the context of the item stem.

- There are two types of constructed-response items: short answer and extended response.
    1. Short-answer items should require a more limited response than extended-response items. A short-answer item may, for example, require students to do one of the following: write a few sentences or equations; complete a table, graph, or chart; draw a picture or construct a diagram; carry out a calculation.
    2. Extended-response items will require students to construct longer responses. For example, students may be required to create a graph showing the appropriate data, title, and labeled axes; create and/or extend tables, diagrams, or pictures; provide a lengthy written explanation or a written explanation that includes equations, pictures, or diagrams.

OVERVIEW OF ITEM SPECIFICATIONS

For each learning target, item specifications are organized under four headings or sections:

- Learning Target
- Item Characteristics
- Stimulus Attributes
- Vocabulary/Mathematical Terms

1. The first heading states the learning target which summarizes one or more benchmarks as identified in the mathematics section of the CSL Essential Academic Learning Requirements Technical Manual, July 17, 1998.
2. Under the second heading are lists of the format the test items may utilize (multiple choice, short answer, and/or extended response) and highlights important points about the concept knowledge/mathematical process being assessed.
3. Under the third heading are lists of attributes of the stimuli (brief written material and/or diagrams, graphs, charts, tables, or drawings) that might precede a question. This section may also note any tools or manipulatives that may be used in answering the question.
4. Under the third heading are lists of mathematical vocabulary or terms that may be used in the items. Other mathematical terms that need explanation or contextual support for students will also be identified.

Note about the Item Specifications:
Some items, particularly constructed-response items, may address several concepts or procedures in a single item. Other items (e.g., multiple choice) may focus on only one concept or procedure.

CONTENT STRANDS AND LEARNING TARGETS

Content Strand 1: Number Sense
NS01 (Numbers and Numeration)

Demonstrates an understanding of and uses the symbolic representations of real numbers; explains the magnitude of numbers by comparing and ordering real numbers. (1.1.1, 1.1.2)

Item Characteristics:

- Multiple-choice and short-answer items can be used to test this learning target.
- Items will assess symbolic representations of real numbers, including the forms of fractions, decimals, percents, integers, positive integer exponents (negative exponents may be used only as part of scientific notation), absolute value, the number line, geometric representations, and pictorial models.
- Test items that ask for understanding of place value by comparing, sequencing, and ordering real numbers will include numbers from the hundred billions place to the sixth decimal place. (Exponents and scientific notation may be used in ordering numbers.)
- In comparing, sequencing, and ordering fractions, students should know how to express fractions in their lowest terms and understand how to convert improper fractions to mixed numbers and vice versa (fractions should be reasonable).
- Students may be asked to illustrate and compare mixed numbers and improper fractions.
- Students should be able to find equivalents between common fractions, decimals, and percents, and between decimals and scientific notation. (Reasonable numbers should be used.)
- Items may have students show or describe relationships between the different forms of real numbers using words or pictures.
- Test items may have students compare, sequence, and order any combination of forms of real numbers.

Stimulus Attributes:

- For whole numbers, test items may include illustrations of number lines and other pictorial models.
- For fractions, decimals, or fractional parts of sets, test items may include illustrations of real-life objects or geometric shapes.

Vocabulary/Mathematical and Terms:

- Terms that can be used: base, common denominator, convert, decimal form, equivalent, exponent, improper fraction, mixed number, negative, numerator, percent, place value, power, scientific notation, simplify
- Terms that cannot be used:

Content Strand 1: Number Sense
NS02 (Number Theory)

Demonstrates an understanding of concepts and processes involving primes, composite numbers, divisibility, and factors and multiples; understands the properties of the system of real numbers. (1.1.1, 1.1.3)

Item Characteristics:

- Multiple-choice and short-answer items can be used to assess this learning target.
- Test items will assess students' understandings of the concepts and processes involving primes, composite numbers, divisibility, and factors and multiples of integers. For example, students may be asked to explain why a number is a prime or a composite number, identify all factor pairs for a given number, tell what numbers are evenly divisible by a given number, or whether a number is a multiple of another number. (Note: The concept of composite numbers should be understood; however the term composite either will be avoided or defined in a stimulus.)
- Both numbers and variables may be used in items assessing this target.
- Items will also require students to demonstrate an understanding of the properties of the system of real numbers. For example, an item may ask students to show or tell why a property is true.
- Properties to be included are:
  Properties of Addition
  - Inverse Property
  - Commutative Property
  - Associative Property
  - Identity Property
  - Zero Property

  Properties of Multiplication
  - Inverse Property
  - Commutative Property
  - Associative Property
  - Identity Property
  - Zero Property

  Distributive Property of Multiplication over Addition
- Items will not require students to name, identify by name, or define properties. Rather, items will ask students to demonstrate or evaluate the applications of these properties.

Stimulus Attributes:

Pictures and charts can be used.

Vocabulary/Mathematical Terms:

- Terms that can be used: prime, divisible, multiple, factor, property, rational, reciprocal, common factor, common multiple, common denominator, opposite
- Terms that cannot be used: the name of a specific property (unless it is defined), greatest common factor, least common multiple

Content Strand 1: Number Sense
NS03 (Conceptual Understanding of Operations)

Demonstrates an understanding of operations on real numbers. (1.1.5)

Item Characteristics:

- Multiple-choice and short-answer items can be used to assess this learning target.
- This learning target involves understanding the underlying concepts and relationships of operations on real numbers.
- Items that assess this target include an understanding of the order of operations (including exponents). Items may ask students to explain the meaning of an operation (but not merely basic operations on whole numbers).
- Items may assess student's conceptual understandings of absolute value (in the context of measurement), exponents, radicals (square roots only), or factorials (if defined).
- Demonstrating an understanding of operations on real numbers includes the relating of multiplication with repeated addition, division with repeated subtraction, addition with subtraction, multiplication with division, and exponents with repeated multiplication.
- Items assessing understanding of the four basic operations should use numbers other than whole numbers (i.e., fractions and negatives).
- Items used to assess this target will be placed only in the "no tools" section of the test.

Stimulus Attributes:

Numbers used in stimulus should be reasonable.

Vocabulary/Mathematical Terms:

- Terms that can be used: absolute value, exponents, numerical expression (which usually includes one or more operations and/or other symbols), order of operations, evaluate, factorial (if defined)
- Terms that cannot be used:

Content Strand 1: Number Sense
NS04 (Estimation)

Identifies situations in which estimation is sufficient and computation is not required; uses estimation to predict computation results and to determine reasonableness of answers. (1.1.8, 1.1.9)

Item Characteristics:

- Multiple-choice and short-answer items can be used to assess this learning target.
- Some of the most commonly used strategies for estimating sums, differences, products, and quotients are listed below, but students may use any reasonable estimation strategy for any items requiring the use of estimation:
  a) front-end estimation: uses front-end digits to obtain an initial estimate, and then makes an adjustment to refine the estimate
  b) rounding: writes a number to the nearest power of ten (An adjustment step is sometimes added to the rounding process.)
  c) compatible numbers: changes or rounds each number in a problem so the number can be manipulated or computed mentally with another number
  d) clustering: rounds numbers around a common value so they can be readily computed mentally
- Students may be asked to evaluate the appropriateness of estimation in different situations, use estimation to predict computation results, or use estimation to determine whether the computation result is reasonable.
- Students will not be asked to do an exact calculation and then round.
- Test items should not require the use of a specific estimation strategy.
- Scoring criteria should consider various estimation strategies. Item development should focus more on the appropriate use of estimation rather than estimation strategies.

Stimulus Attributes:

Charts, tables, diagrams, and illustrations can be used.

Vocabulary/Mathematical Terms:

- Terms that can be used: estimate, estimation, approximate(ly), approximation, reasonableness
- Terms that cannot be used: names of specific strategies

Content Strand 1: Number Sense
NS05 (Computation)

Computes with real numbers, powers, and roots. (1.1.6)

Item Characteristics:

- Multiple-choice and short-answer items can be used to assess this learning target.
- This learning target involves applying the concepts and relationships of operations to real numbers.
- The operations and an understanding of the order of evaluation will be assessed.
- The form of numbers may include percents, decimals (to the place values indicated in target NS01), exponents, radicals (square roots only), scientific notation, and fractions (fractions to be added or subtracted may include those whose least common denominator should be found first, although the fractions should be reasonable).

Stimulus Attributes:

Items may be straightforward computation (including the use of order of operations) or computation may be assessed in the context of a real-life application, such as taxes, interest, or basic accounting.

Vocabulary/Mathematical Terms:

- Terms that can be used: equivalent, evaluate, greatest common factor, least common denominator, mixed number, power, scientific notation, simplify
- Terms that cannot be used:

Content Strand 1: Number Sense
NS06 (Ratio and Proportion)
Demonstrates an understanding and applies concepts of ratio and both direct and inverse variation. (1.1.4)

Item Characteristics:

- Multiple-choice and short-answer items can be used to assess this learning target.
- Items will assess students' understanding of ratio and proportion and the correct application of these concepts in various problem-solving situations.

Stimulus Attributes:

Pictorial representation of numbers, charts, tables, and diagrams may be used in test items.

Vocabulary/Mathematical Terms:

- Terms that can be used: proportion, proportional, rate, ratio, scale, similar (figures), numerator, denominator
- Terms that cannot be used:

Content Strand 2: Measurement
ME01 (Attributes and Dimensions)

Demonstrates an understanding of how changes in dimension affect perimeter, area, and volume; calculates rate and other derived and indirect measurements. (1.2.1, 1.2.3)

Item Characteristics:

- Multiple-choice and short-answer items can be used to test this learning target.
- Items will assess students' understandings of the relationships among linear dimensions, area, and volume. Surface area is included.
- Items may assess students' description and comparison of the impact that changes in measurement have on the measurable attributes of objects and events, such as perimeter, area, and volume. (Other measurable attributes which may be assessed in this target include: monetary units and denominations, temperature, time, mass, capacity, etc.)
- Students will also be required to demonstrate an understanding of rate and to determine and label units (which does not apply only to rate).
- Conversion factors may be used in some problems; however, the units must be within the same system.
- Rates may be expressed either using the slash or the horizontal bar.
- The items should be equally divided between the metric system and the U.S. customary system.

Stimulus Attributes:

Pictorial representations of objects and geometric figures, tables, charts, and graphs may be used. Students are expected to contend with up to two dimensional changes at once.

Vocabulary/Mathematical Terms:

- Terms that can be used: either the U.S. customary or metric systems of measurement and abbreviations common to each system.
- Terms that cannot be used:

Content Strand 2: Measurement
ME02 (Approximation/Precision)

Demonstrates an understanding that the precision and accuracy of measurement is affected by the measurement tools and calculating procedures; uses estimation to obtain reasonable approximations. (1.2.4, 1.2.5)

Item Characteristics:

- Enhanced multiple-choice and short-answer items can be used to assess this learning target.
- Test items will assess students' knowledge of when estimation is appropriate for finding reasonable approximations.
- Items will assess students' estimates and whether they obtain reasonable approximations.
- Items may require students to cross systems for "approximation" only (within reason and for common units, e.g., meters to yards, quarts to liters, etc.).

Stimulus Attributes:

Charts, tables, diagrams, and illustrations can be used.

Vocabulary/Mathematical Terms:

- Terms that can be used: estimate, estimation, approximate, approximation
- Both the U.S. customary and metric units of measurement. (Avoid decimeters, decameters, hectometers, and other uncommon metric units.)
- Terms that cannot be used:

Content Strand 2: Measurement
ME03 (Calculation)

Measures objects and events directly or uses indirect methods; calculates rate and other derived and indirect measurements. (1.2.2, 1.2.3)

Item Characteristics:

- Multiple-choice and short-answer items can be used to assess this learning target.
- Items may assess students' measurements of the dimensions of objects directly if rulers/protractors are available in the classroom or rulers/protractors are provided in the test.
- Items will assess the calculation of perimeter, area, volume, and rates.
- Formulas can be given in the problem.
- The answer key and distractors will be stated in terms of the same system of measurement.

Stimulus Attributes:

Pictorial representations of objects and geometric figures may be used, as well as written descriptions.

Vocabulary/Mathematical Terms:

- Terms that can be used: Test items will include the use of both the U.S. customary and metric systems of measurement and abbreviations common to each system.
- Students are expected to use the following formulas: rectangular area, volume of a rectangular solid, perimeter of polygons, area of triangle, simple rate formulas (D = rt.).

Content Strand 2: Measurement
ME04 (Systems)

Demonstrates an understanding of the benefits of standard units of measurement and the advantages of the metric system; compares, contrasts, and uses both the U.S. and metric systems; selects and uses tools that will provide an appropriate degree of precision. (1.2.6, 1.2.7, 1.2.8)

Item Characteristics:

- Multiple-choice and short-answer items can be used to assess this learning target.
- Test items will assess students' understanding of the benefits of standard and/or nonstandard units of measure as well as appropriate times to use standard and/or nonstandard units.
- Items will assess understanding of the approximate nature of measurement and how the selection and use of tools affects precision and accuracy.
- Test items will assess students' selection and use of appropriate instruments (tools), units (standard or nonstandard), and/or procedures for measuring perimeter, area, volume, and rate.

Stimulus Attributes:

- For items addressing the "selection and use of appropriate instruments" part of this learning target, only instruments that are familiar to tenth-grade students, conceptually or manually, should be considered for use.
- Examples of possible stimuli (either actual or pictorial) include meter stick, ruler, protractor, analog and digital timepieces, thermometer, and scale (including balance).

Vocabulary/Mathematical Terms:

- Terms that can be used: accuracy, accurate
- In test items that address the concepts of nonstandard and standard unit of measurement, illustrations or examples will be included to clarify the meanings of these terms, if they are used.
- Test items will include the use of both the U.S. customary and metric systems of measurement and abbreviations common to each system.
- Terms that cannot be used: precision

Content Strand 3: Geometric Sense
GS01 (Shapes and Figures)

Compares, describes, and classifies 2- and 3-dimensional geometric figures; constructs geometric models and scale drawings using tools as appropriate. (1.3.1, 1.3.2)

Item Characteristics:

- Multiple-choice and short-answer items can be used to assess this learning target.
- Test items will assess students' identification and description of the properties of 2- or 3-dimensional figures and the comparison and classification of the figures according to their properties. For example, a test item may ask students to sort different geometric figures according to those that contain right angles and those that do not.
- For short-answer items, students may be asked to draw one or more geometric figures having certain characteristics or attributes.
- Students may also be asked to translate between 3-dimensional objects and various views of their 2-dimensional representations.

Stimulus Attributes:

Test items will use illustrations of 2-dimensional figures, 3-dimensional figures, and real-life objects when appropriate (e.g., blueprints).

Vocabulary/Mathematical Terms:

- Terms that can be used: face, radius, diameter, circumference, acute, obtuse, vertex/vertices, right angle, 2-dimensional figure, circle, triangle, square, rectangle, rhombus, pentagon, hexagon, octagon, 3-dimensional figure, cube, pyramid, prism, cylinder, cone, sphere, parallelogram, quadrilateral, trapezoid, parallel, perpendicular, plane, similar, congruent, edge
- Terms that cannot be used: space figure, plane figure, solid figure, orthogonal, regular, convex, concave, polyhedron, exterior/interior/vertical angles, parallelepiped, node, skew

Content Strand 3: Geometric Sense
GS02 (Locations)

Demonstrates an understanding of and uses coordinate grids. (1.3.3)

Item Characteristics:

- Multiple-choice and short-answer items can be used to assess this learning target.
- For this outcome, the emphasis will be placed on discussing the location of objects on 2-dimensional coordinate grids.
- Test items will assess students' identification or description of the relative locations of objects to one another or on a 2-dimensional coordinate or location grid (map), using ordered pairs or map coordinates (i.e., letter and number combinations).
- The concept of a 3-dimensional grid may be used only in a real-life context in which the student devises a way to describe a point in space.

Stimulus Attributes:

- Location grids may use ordered number pairs or letter and number combinations.
- Ordered pairs may include negative numbers.

Vocabulary/Mathematical Terms:

- Terms that can be used: (x, y) coordinates, grid, ordered pair, plot, axes, origin, 2-dimensional coordinate grid
- Terms that cannot be used: Cartesian plane, rectangular grid, 3-dimensional coordinate grid

Content Strand 3: Geometric Sense
GS03 (Geometric Properties)

Demonstrates an understanding of simple differences between the geometric properties of a plane and a sphere and uses properties of symmetry, similarity, and congruence. (1.3.4, 1.3.5)

Item Characteristics:

- Multiple-choice and short-answer items can be used to assess this learning target.
- Test items will address simple differences between the geometric properties of a plane (e.g., a "flat Earth") and a sphere (e.g., the real Earth).
- Items will require students to identify, describe, or compare congruent, symmetrical, or similar figures in either illustrations of geometric figures or real-life objects.
- The concept of tessellations may be used to assess the idea of congruent figures used in patterns and real-life objects.
- Short-answer items may have students draw congruent, symmetrical, or similar figures or complete a picture or design given the line of symmetry.
- Other geometric relationships that may be assessed in test items (not necessarily vocabulary) include parallel, intersecting, and perpendicular lines; rays; acute, right, and obtuse angles; isosceles and equilateral triangles; and number of degrees in a circle, triangle, or quadrilateral.

Stimulus Attributes:

- Pictorial representations of geometric figures, as well as illustrations of real-life objects that convey these geometric concepts, will be used in test items.
- Right angles and perpendicular lines may be indicated by a corner box in figures or the ^ symbol.

Vocabulary/Mathematical Terms:

- Terms that may be used: plane, sphere, congruent, similar, symmetrical, line of symmetry, parallel, perpendicular.
- The term tessellation may be used only if an example is provided.
- Terms that cannot be used:

Content Strand 3: Geometric Sense
GS04 (Transformations)

Demonstrates an understanding of and constructs multiple geometric transformations. (1.3.6)

Item Characteristics:

- Multiple-choice and short-answer items can be used to assess this learning target.
- Test items may have students identify a particular transformation or draw a particular transformation of a figure on a grid. The terms are defined as follows:
  a) translation (slide): displacement of a figure that slides the figure a given distance in a given direction
  b) reflection (flip): displacement of a figure about a line, resulting in a "mirror image" of the original figure on the opposite side of the line
  c) rotation (turn): rotational displacement of a figure in which the figure is turned a given angle and direction around a point

Stimulus Attributes:

Geometric transformations that appear in test items may or may not be illustrated on a grid. However, grids will generally be provided in test items that require students to construct a particular geometric transformation.

Vocabulary/Mathematics Terms:

- Terms that can be used: The words translation, reflection, and rotation should be used instead of the words slide, flip, and turn in parentheses, as shown above.
- Terms that cannot be used: displacement, transformation

Content Strand 4: Probability and Statistics
PS01 (Determine Probabilities)

Demonstrates an understanding of the properties of dependent and independent events; understands and uses appropriate counting procedures to determine probabilities; uses both experimental and theoretical methods to determine probabilities. (1.4.1, 1.4.2, 1.4.3)

Item Characteristics:

- Multiple-choice and short-answer items can be used to assess this learning target.
- Students should be able to determine simple experimental probabilities and to discern among events that are sure to happen, events that are sure not to happen, and events that no one can be sure about.
- Students should be able to list all possible outcomes of certain experiments or activities; evaluate or explain why certain outcomes are more (or less) likely to happen than others; and compare or evaluate whether experimental results match or don't match mathematical probability.
- Students should be able to interpret or express the probability of a given event in the form of a ratio or percentage.

Stimulus Attributes:

Pictures of real-life objects (e.g., game spinners, coins, chips, marbles, number cubes), diagrams, charts, tables, graphs, and tree diagrams (if it is partially shown or completed for the student) may be used in items that involve probability.

Vocabulary/Mathematical Terms:

- Terms that can be used: random, probability, population
- Terms that cannot be used: tree diagram, independent, dependent

Content Strand 4: Probability and Statistics
PS02 (Collect and Organize Data) *

Demonstrates an understanding of how to collect data using appropriate methods and technology; organizes and displays data in appropriate forms such as tables, graphs, scatter plots, and box plots. (1.4.5, 1.4.6)
** Actual collection of data will be in the Classroom-Based Evidence Models.*

Item Characteristics:

- Multiple-choice and short-answer items can be used to assess this learning target.
- Test items will assess students' descriptions of a population, and their understandings of how to collect a random sample of data to represent a population.
- Test items may ask students to determine which method of collecting data would be more appropriate for a particular situation.
- Students should understand that data can be represented in many different forms and that there are many ways to go about collecting, organizing, displaying, and interpreting the data.
- Test items may have students identify the kinds of questions that need to be asked in a survey in order to obtain the correct kind of information.
- In evaluating methods for collecting data, students may need to take into account issues such as ensuring random sampling when the situation calls for it; avoiding built-in bias when collecting information; ensuring random sampling when that is what is called for; and not introducing confounding variables.
- Items may require students to make graphs, charts, or tables, and summarize data.
- When the test item focuses on the construction of a table or graph, the item may require students to determine an appropriate label for a graph, or scale for an axis for a bar graph, or to analyze the completeness and accuracy of the data in a table or graph.
- Other items may require students to use the information in the table or graph they constructed to solve an application problem.

Stimulus Attributes:

- Pictures, diagrams, charts, tables, or graphs will be used.
- Kinds of graphs that may be used in these items include pictographs, bar graphs, circle graphs, line graphs, box and whisker graphs, histograms, multiple line graphs, scattergrams, and stem and leaf plots. (Do not use these terms or test vocabulary of these terms.)

Vocabulary/Mathematical Terms:

- The terms survey and random sample will be used in context. Students should understand these terms as related to collecting data or planning to collect data in a systematic way.
- Terms relating to graphs, e.g., scale, bars, and axes, may be used. Students should be familiar with terms such as pattern or trend in analyzing graphs.
- Specific names of graphs can be used in the items. These include pictographs, bar graphs, circle graphs, line graphs, box and whisker graphs, histograms, multiple line graphs, scattergrams, and stem and leaf plots.
- Terms that cannot be used: confounding variables

Content Strand 4: Probability and Statistics
PS03 (Describe and Analyze Data)

Calculates and uses the different measures of central tendency, variability, and range as appropriate in describing a set of data; understands that there can be different interpretations of the same set of data and uses statistics to support different points of view. (1.4.4, 1.4.7)

Item Characteristics:

- Multiple-choice and short-answer items can be used to assess this learning target.
- Test items will focus on real-life situations.
- Items may include data presented as pictographs, time lines, bar graphs, line graphs, circle graphs, line plots, box and whisker graphs, and tables.
- Items may assess students' descriptions, interpretations, and evaluations of other interpretations of data to support arguments.
- Items may require students to make graphs or tables, note trends and patterns, and summarize and analyze data, including calculating and using mean, median, mode, and range as appropriate to describe the data.

Stimulus Attributes:

Pictorial representations (e.g., icons), graphs, tables, and charts will be used in test items.

Vocabulary/Mathematical Terms:

- Terms that can be used: Terms relating to graphs, e.g., scale, grid, bars, axes, and plot may be used.
- Students should be familiar with terms such as pattern or trend.
- Terms that cannot be used: bias, quartile (may assess the concept, but do not use the term)

Content Strand 4: Probability and Statistics
PS04 (Make Inferences and Predictions)

Designs and conducts experiments to verify or disprove predictions; understands and makes inferences based on analysis or experimental results. (1.4.8, 1.4.9)

Item Characteristics:

- Multiple-choice and short-answer items can be used to assess this learning target.
- Students should be able to find characteristics of large groups by randomly selecting a small group.
- Test items may assess a students' making of inferences based on the analysis of experimental results and graphical representations.
- Items can include interpretation of percentage of error or margin of error (± __%).

Stimulus Attributes:

- Pictures, diagrams, charts, tables, or graphs will be used.
- Kinds of graphs that may be used in these items include pictographs, bar graphs, circle graphs, line graphs, box and whisker graphs, histograms, multiple line graphs, scattergrams, and stem and leaf plots. (Do not use these terms or test vocabulary of these terms.)

Vocabulary/Mathematical Terms:

- Terms that can be used: predict, ratio, percent, combinations (use only if appropriate and mathematically correct), and probability, as well as the phrase "What are the chances of ....?"
- Term that cannot be used: permutations (use this phrase instead: "How many different ways ....?")

Content Strand 5: Algebraic Sense
AS01 (Patterns, Series, and Sequences)

Recognizes, creates, extends, and generalizes patterns, series, and sequences. (1.5.1)

Item Characteristics:

- Multiple-choice and short-answer items can be used to assess this learning target.
- Test items will assess students' recognition, creation, and extension of a variety of patterns, and the use of patterns to make generalizations and predictions.
- Items may require students to continue a pattern by identifying or supplying missing element(s) and/or describing the rule for the pattern.
- Number patterns will be emphasized more than pictorial (visual or geometric) patterns, but both will be represented.
- Number patterns may be presented in a sequence or in a table of number pairs. (The number pattern should follow a rule.)
- A number pattern may be a combination of one or two operations (addition, subtraction, multiplication, division, or simple exponents).

Stimulus Attributes:

Charts, tables, diagrams, and pictorial representations of objects or geometric shapes may be used.

Vocabulary/Mathematical Terms:

- Terms that can be used: rule and sequence, with regard to explaining or finding a pattern.
- Terms that cannot be used: geometric sequence/series, arithmetic sequence/series

Content Strand 5: Algebraic Sense
AS02 (Represent Functions and Relationships)

Understands, develops, and expresses rules for patterns; translates among tabular, symbolic, and graphical representations of relations. (1.5.2, 1.5.3)

Item Characteristics:

- Multiple-choice and short-answer items will be used to assess this learning target.
- Test items will assess students' representation (using words and/or symbols) and extension of number patterns in tables or graphs using words or symbols. The number pattern may involve one or two operations.
- Items will also assess students' translations among tabular, symbolic, and graphical representations of relations; their identification of correct versus incorrect representations of reasonable situations; their creation of situations that match representations and vice versa.
- The representations will include variable quantities and may be expressions, equations, or inequalities.

Stimulus Attributes:

Charts, tables, lists, and symbolic or written rules showing number patterns may be used. (Students may be asked to use symbols used in computer program/spreadsheet formulas, if the symbols they will need are defined.)

Vocabulary/Mathematical Terms:

- Terms that can be used: rule and sequence (with regard to explaining or finding a pattern), expression, relationship, unknown, variable, algebra, algebraic, table
- Terms that cannot be used: tabular

Content Strand 5: Algebraic Sense
AS03 (Equations)

Creates and solves equations and inequalities; evaluates and simplifies expressions. (1.5.5, 1.5.6)

Item Characteristics:

- Multiple-choice and short-answer items can be used to assess this learning target.
- Test items may require students to simplify expressions or solve equations (including systems) or inequalities (including applying the distributive property), and to evaluate by substitution in a problem-solving context.

Stimulus Attributes:

Pictorial representations, charts, tables, graphs, equations, or inequalities may be used in test items.

Vocabulary/Mathematical Terms:

- Terms that can be used: expression, evaluate, set up (verb), relationship, simplify, unknown, value, variable
- Absolute value may be used in an equation, but not in an inequality.
- Also, negatives should not be used in inequalities.
- Terms that cannot be used: quadratic, setup (noun)

Content Strand 5: Algebraic Sense
AS04 (Application of Equations)

Represents situations that involve variable quantities with expressions, formulas and equations, and inequalities. (1.5.4)

Item Characteristics:

- Multiple-choice and short-answer items can be used to assess this learning target.
- Items will assess students' identification of correct and incorrect representations of reasonable real-life situations, description of situations to match representations, or creation of representations to match situations.

Stimulus Attributes:

Equations, graphs, charts, and tables in a problem-solving context may be used.

Vocabulary/Mathematical Terms:

- Terms that can be used: model (as a noun only for reference to, not requiring the student to build a model), relationship, value.
- Terms that cannot be used: linear, exponential

Process Strand 6: Solving Problems
SP01 (Investigates Situations)

Analyzes and uses multiple strategies; identifies missing/extraneous information and compensates for it; and searches systematically for patterns in complex situations. (2.1.1, 2.1.2, 2.1.3)

Item Characteristics:

- Multiple-choice, short-answer, and extended-response items can be used to test this learning target.
- Items will be constructed that ask students to identify, use, evaluate, modify, or create methods that can be used to investigate problems or conduct explorations.
- Students may be asked to identify missing or extraneous information or identify patterns that might be helpful in solving a problem.
- Whenever possible, open-ended items will be constructed to allow for more than one approach, and some open-ended items may have more than one correct answer.
- The focus is on the method rather than on the solution (which is part of SP03).

Note: Students will not be expected to know problem-solving methods by name.

Stimulus Attributes:

Mathematical situations may be modeled with word problems, illustrations, and/or other graphic materials. The use of manipulatives may be incorporated into constructed-response items.

Vocabulary/Mathematical Terms:*

- Terms that can be used: relevant, process
- Terms that cannot be used: extraneous, irrelevant

** Since stimuli and situations involved in process strands will be similar to those used in concepts and procedures, students should be familiar with the terms noted previously. This holds true for all process targets.*

Process Strand 6: Solving Problems
SP02 (Defines the Problem)

Defines the problem to be solved in complex situations; and identifies unknowns and questions to be answered in complex situations. (2.2.1, 2.2.2, 2.2.3)

Item Characteristics:

- Multiple-choice, short-answer, and extended-response items can be used to test this learning target.
- Items will be constructed to assess students' understandings of conditions, variables, subproblems, patterns, questions to be answered, and unknown data in problem-solving situations.
- Given a problem, students could be asked to:
  1. define or re-state problems in their own words
  2. identify unknowns or questions that need to be answered in finding a solution; identify patterns or data that are useful or necessary in understanding and solving a problem
  3. define or formulate questions or sub-problems that could be used in a multi-step problem or process.
- Items may also ask students to formulate problems themselves. For example, they may be asked to write variations for problems previously explored; and to identify questions that could be answered by investigating the data from a chart, table, menu, advertisement, or other representations of data.
- Students should understand that sometimes there are "smaller" problems that need to be solved in the process of solving a larger problem

Stimulus Attributes:

- Mathematical situations may be modeled with word problems, illustrations, and/or other graphic materials.
- Charts, tables, menus, advertisements, or other sources of data may be used.

Vocabulary/Mathematical Terms:

- Terms that can be used: See all content strands for specifics.
- Terms that cannot be used: Names of any specific problem-solving processes; understanding of that term "sub-problem" will not be tested

Process Strand 6: Solving Problems
SP03 (Constructs Solutions)

Organizes and synthesizes information from multiple sources; and selects and uses appropriate tools, methods, operations, and processes to construct solutions. (2.3.1, 2.3.2, 2.3.3)

Item Characteristics:

- Multiple-choice, short-answer, and extended-response items can be used to test this learning target.
- Given a problem situation and some display of data (in charts, tables, graphs, diagrams, text, etc.), students could be asked to collect and organize all relevant or needed information and then apply various concepts, procedures, and strategies to construct a solution.
- Given a problem situation and tools, or a choice of tools, students could be asked to select among the tools for effectiveness, and use the tool(s), problem-solving strategies, and/or various concepts and procedures in constructing solutions.

Stimulus Attributes:

Charts, tables, graphs, pictures, and other representations of data are likely material for stimuli.

Vocabulary/Mathematical Terms:

- Terms that can be used: See content strands for specifics.
- Terms that cannot be used: See specific content strand limitations.

Process Strand 7: Reasoning Logically
RL01 (Analyzes Information)

Interprets and integrates information from multiple sources. (3.1.1)

Item Characteristics:

- Multiple-choice, short-answer, and extended-response items can be used to test this learning target.
- Items will assess students' interpretation, integration, comparison, and contrast of data from several sources, in terms of the following:
    1. reliability of data
    2. knowledge of when and how data may be combined
    3. patterns shown by data
    4. agreements or discrepancies between data
    5. reasons for conflicting or varying data
    6. conclusions that are supported or refuted by data.

Stimulus Attributes:

Charts, tables, maps, graphs, diagrams, pictures (both two- and three-dimensional), and other representations of data or information are likely material for stimuli.

Vocabulary/Mathematical Terms:

- Terms that can be used: compare, contrast, analysis/analyze
- Terms that cannot be used: See specific content strands limitations.

Process Strand 7: Reasoning Logically
RL02 (Predicts)

Makes and explains conjectures and inferences based on analysis of problem situations. (3.2.1)

Item Characteristics:

- Multiple-choice, short-answer, and extended-response items can be used to test this learning target.
- Items will assess students' making and supporting conjectures, inferences, predictions, and conclusions based on their analyses of a problem/situation.
- Students may be asked to use a given partial or complete analysis to make conjectures, inferences, predictions, and conclusions.

Stimulus Attributes:

Problems/situations may include maps, technical information or "results" in various formats (charts, tables, graphs), or pictures or other collections/representations of information.

Vocabulary/Mathematical Terms:

- Terms that can be used: Analysis/analyze, conclusion, support, representation, conjecture, inference. Also, see specific content strands.
- Terms that cannot be used: See specific content strand limitations.

Process Strand 7: Reasoning Logically
RL03 (Verifies)

Validates thinking and mathematical ideas using models, known facts, patterns, relationships, counter-examples, and proportional reasoning; tests conjectures and inferences by formulating a proof or by constructing a counter-example; supports arguments and justifies results using inductive and deductive reasoning; checks for reasonableness of results; evaluates and reflects on procedures and results and makes necessary revisions. (3.1.2, 3.3.1, 3.3.2, 3.3.3, 3.3.4)

Item Characteristics:

- Multiple-choice, short-answer, and extended-response items can be used to test this learning target.
- Students may be given examples of :
  1. mathematical thinking
  2. procedures
  3. problem situations
  4. results
  5. conjectures and inferences

  and be asked to:
  1. identify or use known facts, patterns, relationships, models, concrete or abstract examples that contradict a mathematical argument
  2. support proportional reasoning
  3. use deductive or inductive reasoning to evaluate procedures or results, validate positions, or provide explanations.

Stimulus Attributes:

Mathematical and pictorial models, patterns, charts, flow charts, tables, graphs, examples/counterexamples, and other informational sources are likely material for stimuli.

Vocabulary/Mathematical Terms:

- Terms that can be used: evaluate, valid, logic, logical, model (noun), simplify
- Terms that cannot be used: conjecture, position (OK if it refers to location), deduce, argument, proof, validate, verify
- Students should have had practice in using deductive and inductive reasoning, though understanding of those terms as such will not be tested.

Process Strand 8: Communicating Understanding
CU01 (Gathering Information)

Develops or selects an efficient system for collecting information; uses reading, listening, and observation skills to access and extract mathematical information from multiple sources. (4.1.1, 4.1.2)

Item Characteristics:

- Multiple-choice, short-answer, and extended-response items can be used to test this learning target.
- Items will assess students' creation of a plan for collecting information, according to the information desired or the conditions stipulated.
- Students will be asked to:
  1. diagram or describe their general procedure/order of steps
  2. describe or explain the kinds of questions they might ask in gathering information, and their reasons for those questions
  3. recognize the difference between gathering useful information and information that will not be useful, including the following: collecting exactly the information that is sought, not "variations" on that information.
- Items will assess how students' obtain mathematical information from various sources (pictures, diagrams, models, text, symbolic representations, and technology).
- Students may be asked to gather a particular kind of information (according to what is desired) from various sources.

Stimulus Attributes:

- Stimuli will detail any conditions/criteria that a plan should meet. Symbolic and pictorial representations of mathematical information, charts, tables, graphs, diagrams, models, and applicable technology are likely materials for stimuli.
- Manipulatives may be incorporated into constructed-response items.

Vocabulary/Mathematical Terms:

- Terms that can be used: variable (in the experimental sense), random, sample
- Terms that cannot be used: See content strands for limitations.

Process Strand 8: Communicating Understanding
CU02 (Organizing and Interpreting & Representing and Sharing)

Organizes, clarifies, and refines mathematical information in multiple ways; expresses complex ideas and situations using mathematical language and notation in appropriate and efficient forms; clearly and effectively expresses/presents mathematical ideas using both everyday and mathematical language which is appropriate to the audience. (4.2.1, 4.3.1, 4.3.2)

Item Characteristics:

- Multiple-choice, short-answer, and extended-response items can be used to test this learning target.
- Students may be asked to take unorganized information, including text, and to
  1. organize it according to certain directions
  2. clarify it; or
  3. represent such information and mathematical understanding in different forms (including tables, pictures, graphs, schematics, algebraic notation, and text) or to translate between forms
- Students may be asked to explain
  1. their understanding of mathematical ideas or presentations
  2. mathematical ideas and understandings
  3. in ways that would be appropriate to audiences the students might normally be in contact with (younger students, students of the same age, speakers of a different language, parents).
- Students may be asked to provide written explanations of mathematical representations that are largely pictorial or schematic; to translate from symbols into text or translate a diagram or physical model into symbols or words.

Stimulus Attributes:

Mathematical information or representations will be the focus of stimuli.

Vocabulary/Mathematical Terms:

- Terms that can be used: mathematical language, audience
- Terms that cannot be used: graphic

Process Strand 9: Making Connections
MC01 (Connections among Concepts and Procedures and between Different Mathematical Representations)

Relates and uses conceptual and procedural understandings among multiple mathematical content areas or relates and uses multiple equivalent mathematical models and representations. (5.1.1, 5.1.2)

Item Characteristics:

- Multiple-choice, short-answer, and extended-response items can be used to test this learning target.
- Given a procedure, concept, or problem, students may be asked to describe, demonstrate, or apply relationships among procedures or concepts. These may be within a conceptual strand or across two or more strands. Subject areas are: number sense, measurement, geometric sense, probability and statistics, and algebraic sense.
- Given a mathematical representation or problem, students may be asked to identify equivalent (or partly equivalent) representations, to determine how (or to what extent) they are equivalent
- Given a mathematical representation or problem, students may be asked to create representations that are equivalent to given ones
- Given a mathematical representation or problem, students may be asked to use or create representations to make or show connections within mathematical situations.
- Mathematical representations may be verbal, pictorial, numerical, geometric, graphical, or algebraic.

Stimulus Attributes:

Mathematical information or representations will be the focus of stimuli. Pictorial and other representations of data, verbal, numerical, geometric, graphical, algebraic, or other mathematical representations are material for stimuli.

Vocabulary/Mathematical Terms:

- Terms that can be used: relationship; see content strands for specifics
- Term that cannot be used: translate; see content strands for specific limitations

APPENDIX C

WASHINGTON ASSESSMENT OF STUDENT LEARNING

GRADE 10

READING

TEST SPECIFICATIONS

Washington Assessment of Student Learning
Test Specifications
Grade 10 Reading
August 2000

I. PURPOSE

The purpose of this test is to measure Washington tenth-grade students' level of proficiency in the Essential Academic Learning Requirements in reading. The reading test will contain literary, informational, and task-oriented reading selections. All reading selections, up to 800 words and written at a difficulty level appropriate to their audience, will be accompanied by test items that assess proficiency in the components of the Essential Academic Learning Requirements in reading. Test items will be of the following types in the proportions shown:

- Multiple-choice: Student chooses from four responses provided (26-30 items)
- Short-answer: Student constructs short response–phrase(s), sentence(s), or paragraph(s) (9-11 items)
- Extended-response: Student constructs longer, more sustained response–sentences or paragraph(s) (2 items)

Each reading test form will attempt to test all identified Learning Targets (this term has been adopted for the sake of consistency between the state-level tests in the different subject areas) for each Strand, but this goal may not always be practical; not every text will allow every type of question to be asked. The test will offer three subscale reports in Reading for Literary Experience and three in Reading to Learn New Information and Reading to Perform a Task. (These reports correspond to the six Strands shown below.)

The first report for each type of text will reflect students' comprehension of important ideas and details; the second will reflect students' ability to analyze, interpret, and synthesize; the third will reflect students' ability to think critically about what they read. (In the chart below, EALR numbers are given in parentheses after each Learning Target.)

II. STRANDS AND LEARNING TARGETS

Reading for Literary Experience:

Strand LC: Comprehends important ideas and details

Given a literary text to read silently, learners respond to items in which they use textual evidence to

1. Demonstrate understanding of theme or message and supportive details (2.1.2)
2. Summarize text (2.1.2)
3. Make inferences or predictions (2.1.4)
4. Interpret general and specialized vocabulary critical to the meaning of the text (1.2.1)

Strand LA: Analyzes, interprets, and synthesizes

Given a literary text to read silently, learners respond to items in which they use textual evidence to

5. Apply understanding of literary elements (genres; story elements such as plot, character, setting; stylistic devices) and graphic elements/illustrations (1.4.3 or 1.4.1)
6. Compare/contrast elements of the text or make connections within the text (2.2.1)
7. Compare/contrast or make connections between or among texts or synthesize information from a variety of resources (2.2.1)

Strand LT: Thinks critically

Given a literary text to read silently, learners respond to items in which they use textual evidence to

8. Analyze author's purpose and evaluate effectiveness for different audiences (includes fact/opinion, author's point of view, tone, and use of persuasive devices, bias, stereotype) (2.3.2, 2.3.3)
9. Evaluate reasoning and ideas/themes related to the text (2.3.1 or 2.3.4)
10. Extend information beyond text (make generalizations beyond the text to a broader idea or concept, or apply information to other texts or situations, or give a response to reading) (2.3.1 or 2.3.5, 2.3.6, 2.3.7)

Reading to Learn New Information and Reading to Perform Tasks

Strand IC: Comprehends important ideas and details

Given an informational or task-oriented text to read silently, learners respond to items in which they use textual evidence to

11. Demonstrate understanding of major ideas and supportive details (2.1.2)
12. Summarize text (2.1.2)
13. Make inferences or predictions (2.1.4)
14. Interpret general and specialized vocabulary critical to the meaning of the text (1.2.1)

Strand IA: Analyzes, interprets, and synthesizes

Given an informational or task-oriented text to read silently, learners respond to items in which they use textual evidence to

15. Apply understanding of text features (titles, headings, and other information divisions, table of contents, indexes, glossaries, prefaces, appendices, captions) and graphic features (1.5.2)

16. Compare/contrast elements of the text or make connections within the text (2.2.1)

17. Compare/contrast or make connections between or among texts or synthesize information from a variety of resources (2.2.1)

Strand IT: Thinks critically

Given an informational or task-oriented text to read silently, learners respond to items in which they use textual evidence to

18. Analyze author's purpose and evaluate effectiveness for different audiences (includes fact/opinion, author's point of view, tone, and use of persuasive devices, bias, stereotype) (2.3.2, 2.3.3)

19. Evaluate reasoning and ideas related to the text (2.3.1 or 2.3.4)

20. Extend information beyond text (make generalizations beyond the text to a broader idea or concept, or apply information to other texts or situations, or give a response to reading) (2.3.1 or 2.3.5, 2.3.6, 2.3.7)

III. READING PASSAGES

Reading passages used in the test should be drawn from published sources.

- Literary text may include poetry, essays, short stories, novel excerpts, biographies, nonfiction narratives, and plays.
- Informational text may include encyclopedias and other reference materials, pamphlets, book excerpts, newspaper and magazine articles, letters to an editor.
- Task-oriented text may consist of schedules, maps, recipes, instructions, newspaper want ads, consumer reports, travel books, first aid or other manuals, catalogs, yellow pages, credit card or job applications, and other such pieces likely to be within the experience of a tenth-grade student.

1. Where possible, selections should promote a school-to-work connection. Where appropriate, the passages should utilize illustrations and other graphic features.
2. Each assessment will contain one or more selections made up of two or more short passages, e.g., a poem and a short piece of fiction, or a set of directions, a short informational text, and a short essay; these groupings will allow construction of items that call for students to make connections among texts.
3. Passages should not exhibit cultural or other forms of bias that might disadvantage any group (or groups) of students and should avoid subject matter that might prompt emotional distress on the part of some students.
4. It is critical that the reading texts used be well written, of interest to tenth-grade students, and, in all appropriate cases, factually correct.
5. Reading test passages should also reflect Washington's cultural diversity, and as they are presented they should possess structural integrity that allows them to be self-contained.
6. Permission to use selections from copyrighted material will be obtained as necessary.

7. The reading difficulty of the passages will be validated using traditional readability formulas and teacher judgment.
8. The passages' overall suitability will be judged by the committee.

IV. TEST ITEMS

1. Reading test items should always focus on what is essential and consequential in a given text.
2. Test items should be tied closely and particularly to the passage from which they derive, so that the impact of outside knowledge, while never wholly avoidable, can be diminished.
3. Test items should be varied and should address as many strands/learning targets as the passages allow.
4. The test will contain 26-30 multiple-choice items; each of these will have one correct answer and three distractors, the four choices being approximately the same length, having the same format, and being syntactically and semantically parallel.
5. The test will contain 9-11 short-answer items; each of these will require students to construct a short response, defined as phrase(s), sentence(s), or a paragraph, and will focus on one learning target.
6. The test will also contain 2 extended-response items; each of these will require students to construct a longer, more sustained response, defined as sentences or paragraph(s), and will focus on one learning target.
7. Order of presentation of item types will be dictated by logic.
8. With grouped passages, items for each will follow each passage, or they may follow two passages together that should not be separated. Items that require connections between and among the passages will appear last.
9. There should be more than two items that require students to make connections between and among the passages.

V. DISTRIBUTION OF READING SELECTIONS AND TEST ITEMS

Each reading test form will include three literary selections, generating approximately half the total test points, and one or two informational selections and one or two task-oriented selections, generating approximately half the total test points. In addition, one or more selections in each form of the test will consist of two or more short passages, e.g., a poem and a short piece of fiction, or a set of directions, a short informational text, and a short essay. These groupings will allow construction of items that call for students to make connections between and among texts.

- Many of the selections will be short, e.g., 300 words.
- One selection on a form will be longer (as long as 800 words) to allow for development of items that go with more extended text.
- The reading selections together total about 2500 words.
- Total number of multiple-choice items is not to exceed 30.
- Total number of short-answer items is not to exceed 11.
- Total number of extended-response items is not to exceed 2.

The following table contains a matrix showing item distribution by text types, strands, and learning targets in the test.

Reading Test Map for Grade 10 WASL

Text types/Strands	Number of Learning Targets	No. of Reading Selections	No. of MC Items	No. of SA Items	No. of ER Items	No. of Words Per Passage
Literary Selections	10	3	10-15	3-6	1	up to 1300
Comprehends important ideas and details	4		3-5	1-2		
Analyzes, interprets and synthesizes †	3		2-5	1-3	0-1	
Thinks critically	3		2-5	1-3	0-1	
Informational and Task-oriented Selections	10	3-4	10-15	3-6	1	up to 1300
Comprehends important ideas and details	4		3-5	1-2		
Analyzes, interprets and synthesizes †	3		2-5	1-3	0-1	
Thinks critically	3		2-5	1-3	0-1	
Totals	20	6-7	26-30	9-11	2	up to 2500

VI. TEST AND ITEM SCORING

Each multiple-choice item is worth 1 point, each short-answer item is worth 2 points, and each extended-response item is worth 4 points.

Reading Test: Typical distribution of score points by item type

Type	Number of Items	Total Points	Percent of the Total Score
Multiple-choice	28	28	50%
Short-answers	10	20	36%
Extended-response	2	8	14%
Total	40	56	

VII. GENERAL CONSIDERATIONS

1. It is not possible to measure every learning target on every form of the test. However, learning targets from each strand must be tested on each form.
2. The material presented will be culturally diverse, well written, and of interest to tenth-grade students, and the passages and items will be fairly presented in order to gain a true picture of students' reading skills.
3. Each multiple-choice item will contain a question (or incomplete statement) and four answer (or completion) options, only one of which is correct. Correct answers will be distributed approximately evenly among A's, B's, C's, and D's.
4. Each short-answer or extended-response item will give clear indications of what is required of students; a scoring guide will be developed for each constructed-response item, and information from the pilot testing will be used to refine these scoring guides for use with the final forms of the test. Scoring guides will follow a "focused holistic" model in which the score for the response is based on overall quality but also results from a focus on several important features of the student's performance.

- Short-answer items will be scored with a 3-level scoring guide (0-2) in which students may receive full credit, partial credit, or no credit.
- Extended-response items will be scored with a 5-level scoring guide (0-4); the levels may be summarized as Extensive, Essential, Partial, Unsatisfactory, and No credit.

5. To the greatest extent possible, no item on the test will "clue" the answer to any other item.
6. The strand and learning target assessed will be specified for each item.
7. Items will not display unfair representations of genders, races, disabled individuals, or cultural or religious groups. Items will not contain elements that may disadvantage a particular gender, race, culture, religious, or disabled group.
8. Across all forms, there will be balance of gender and active/passive roles by gender.
9. No resource materials may be used by students during testing.
10. Responses will be scored with emphasis on communication of ideas. Conventions of writing (sentence structure, word choice, usage, grammar, and mechanics) will be disregarded unless they seriously interfere with communication.

Estimated Time for Reading Test: 75 minutes, broken into two sessions.

VIII. Addendum (Feb. 1999) SCORING OF OPEN-ENDED ITEMS

Individual scoring criteria will be developed for each constructed-response item. Short-answer items will be scored on a scale of 0 to 2 points, and extended-response items will be scored on a scale of 0 to 4 points. The following scoring criteria are used to assess basic comprehension of main ideas and details and analysis, interpretation, and critical thinking about text. Specific scoring criteria will be developed for each item based on these generic rules.

Scoring Rules for Short Answer Items

Scoring rules for items that assess <u>main ideas and details</u>:

2 A two point response:
- shows thorough comprehension of main idea and important details
- uses ample, relevant information from text(s) to support responses

1 A one point response:
- shows partial comprehension of main idea and important details (may grasp main idea but show difficulty distinguishing between important and unimportant details; may miss part of fundamental who/what/where/when/why)
- attempts to use information from text(s) to support responses; support may be limited or irrelevant

0 A 0 point response shows little or no understanding of the passage main ideas and details.

Scoring rules for items that assess <u>analysis, interpretation, and critical thinking about text</u>:

2 A two point response:
- analyzes appropriate information and/or makes thoughtful connections between whole texts/parts of texts
- develops thoughtful interpretations of text
- uses sufficient, relevant evidence from text(s) to support claims

1 A one point response:
- analyzes limited information and/or makes superficial connections between whole texts/parts of texts
- develops conventional or simplistic interpretations of text
- attempts to use evidence from text(s) to support claims; support may be limited or irrelevant

0 A 0 point response shows little or no understanding of the passage main ideas and details.

Scoring rules for items that assess <u>summarizing and paraphrasing main ideas</u>:

2 A two point response shows thorough comprehension of main ideas

1 A one point response shows partial comprehension of main ideas

0 A 0 point response shows little or no understanding of the passage main ideas and details.

Scoring Rules for Extended Response Items

Scoring rules for items that assess <u>analysis, interpretation, and thinking about text</u>:

<u>4 Points: Meets all relevant criteria</u>

- thoroughly analyzes appropriate information and/or makes insightful connections between whole texts/parts of texts
- develops insightful interpretations of text
- uses ample, relevant evidence from text(s) to support claims

<u>3 Points: Meets most relevant criteria</u>

- analyzes appropriate information and/or makes thoughtful connections between whole texts/parts of texts
- develops thoughtful interpretations of text
- uses sufficient, relevant evidence from text(s) to support claims

<u>2 Points: Meets some relevant criteria</u>

- analyzes limited information and/or makes superficial connections between whole texts/parts of texts
- develops conventional or simplistic interpretations of text
- attempts to use evidence from text(s) to support claims; support may be limited or irrelevant

<u>1 Point: Meets few relevant criteria</u>

- shows difficulty analyzing information and/or makes weak connections between whole texts/parts of texts
- may not develop beyond literal interpretation of text
- uses little or no evidence to support claims

Washington Assessment of Student Learning
Item Specifications
Grade 10 Reading
August 2000

The purpose of this test is to measure Washington tenth-grade students' level of proficiency in the Essential Academic Learning Requirements in reading. To thoughtfully and equitably achieve this goal, test items will be developed that ask students to demonstrate that they can comprehend important ideas and details in a text and that they can analyze, interpret, and think critically about a text. Given the reading test's goal, test items must be reliable and fair and address the learning targets listed in the following pages.

There are some general considerations that can make the task of item writing more efficient and fruitful. These considerations include, but are not limited to, the following:

- Ask questions that deal with issues and details that are of consequence in the selection and central to students' understanding and interpretation of a given text.
- Take care that all items avoid language that shows bias or is otherwise likely to be offensive to or to disadvantage a particular group of students.
- Word items precisely and clearly. The better focused an item, the more reliable and fair it is certain to be, and the more likely all students will understand in the same way what is required of them.
- Make sure that all multiple-choice options–key and distractors–are similar in length and in syntax; students should not be able to rule out a wrong answer or identify a correct response simply by virtue of its looking or sounding different.
- Create distractors that adopt the language and sense of material in the selection so that students must think their way to the correct answer rather than simply identify incorrect responses by virtue of a distractor's obviously inappropriate nature.
- Distractors should always be plausible (but of course incorrect) in the context of the selection.
- All items must clearly indicate what is expected in a response and must help students focus their responses.
- Constructed-response items are of two types: short-answer and extended-response.
- Short-answer items should require a more limited response than extended-response items.
- The two types of constructed-response items may also be differentiated by the number of lines available for the response.

In summary: Reading test items should ask questions that address issues of importance in a text, and those questions should be consequential, concise, focused, and fair.

Additional Rules for Developing Constructed-response Items: Grade 10

Short Answer Item Rules

- Items should ask students to give evidence they have understood or analyzed text. This usually means ending with "Support your answer with information from the story" or, preferably, "Support your answer with an example (or two examples, details, ideas) from the story."
- Item stems will be clear and succinct as possible.
- Each short-answer item will give clear indications of what is required of students (e.g., if two examples are required, the stem will indicate this; if a generalization and a supporting example are required, the stem will indicate this).
- Anything required by the scoring rule will be asked for in the item stem.
- Any short-answer item that requires personal response or connections beyond the text will ask for references to how the student's ideas relate to the text (e.g., "How did the story help you . . .?").
- Do not use the format that gives students a choice of (a number of) viable answers and require students to choose one and support it.
- As much as possible, the language of the items should reflect the language of the strands and learning targets (and thus the Essential Academic Learning Requirements). So, for example, ask students to "analyze" when appropriate.
- Items should not be broken into parts with lines following each part.

Extended Response Item Rules

- Items should ask students to give evidence they have understood or analyzed text. This usually means ending with "Support your answer with information from the story" or, preferably, "Support your answer with three specific examples (details, ideas) from the story."
- Item stems will be clear and succinct as possible.
- Each extended-response item will give clear indications of what is required of students (e.g., if a generalization and three supporting examples are required, the stem will indicate this).
- Anything required by the scoring rule will be asked for in the item stem.
- Any extended-response item that requires personal response or connections beyond the text will ask for references to how the student's ideas relate to the text (e.g., "How did the story help you . . .?"). NOTE: AVOID PERSONAL RESPONSE EXTENDED.
- Do not use the format that gives students a choice of (a number of) viable answers and require students to choose one and support it.
- As much as possible, the language of the items should reflect the language of the strands and learning targets (and thus the Essential Academic Learning Requirements). So, for example, ask students to "analyze" when appropriate.
- Items should not be broken into parts with lines following each part.

For each learning target, item specifications are organized under three headings or sections:

- Learning Target
- Stimulus Attributes
- Item Description

- The first heading states the learning target, which summarizes one or more benchmarks as identified in the reading section of the Essential Academic Learning Requirements (EALRs) Technical Manual (February 26, 1997).
- The second section indicates the appropriate characteristics (format and content) of the reading stimuli that will precede any items.
- The third section gives a description of the items that can be used to assess the learning target. This will include the rules for item types, item formats, response formats, and distractors.

Strand LC: Comprehends important ideas and details

Learning Target 1: Demonstrate understanding of theme or message and supportive details. (2.1.2)

Stimulus Attributes:

1. Format: Literary text
2. Content: The selection must contain a clear theme or message that is supported by details.

Item Description:

1. Possible item type: M, S
2. Item format: Clearly states requirement that students demonstrate understanding of a theme or message or of the details that support it.
3. Response format: Response may be phrase(s), sentence(s), or paragraph(s).
4. Multiple-choice items:

    Example of stem:

    What is the author's message in this story?

    Example of response choices:

    a. The correct response is the best statement of the message of the story.
    b. An incorrect response may contain an idea not included in the passage.
    c. An incorrect response may contain an unimportant idea that is stated in the passage.

    Example of stem:

    The message of this poem has to do with sharing. Which of the following ideas from the poem helps get this message across?

    Example of response choices:

    a. The correct response is a detail that supports the message.
    b. An incorrect response may be a detail in the poem that does not support the message.
    c. An incorrect response may be a detail that appears to support the message but is not actually included in the selection.

5. Short-answer items:

    Example of item:

    What is the theme of this poem? Give a specific example from the poem to support your answer.

Strands LC/IC: Comprehends important ideas and details

Learning Targets 2 and 12: Summarize text. (2.1.2)

Stimulus Attributes:

1. Format: Literary, informational, task-oriented text

2. Content: The selection must be substantive enough to allow for the type of summary called for in the item.

Item Description:

1. Possible item type: M, S

2. Item format: Clearly states requirement that students choose or construct a reasonable summary of a text.

Note: The content committee prefers that students write rather than choose summaries. This cannot be done across the board, because the number of short answer items is limited for each test. However, it could be done at least once per test. So the short-answer option below should be used with about a third or a fourth of the passages. Then, in operational forms of the test, we can attempt to include one written summary each year.

3. Response format: Response may be phrase(s), as in a title, sentence(s), or paragraph(s).

4. Multiple-choice items:

Example of stem:

Which sentence best summarizes this (selection)?

Example of response choices:

a. The correct response is the best summary.
b. An incorrect response may contain a related idea not included in the passage.
c. An incorrect response may contain an idea from the passage that is too narrow to be acceptable as a summary.

5. Short-answer items:

Example of item: (informational)

In your own words, write a summary of the main points of this selection. Be sure to include at least (number) points in your summary.

Example of item: (literary)

In your own words, write a summary of the main events in this story. Be sure to include at least (number) points in your summary.

Strands LC/IC: Comprehends important ideas and details

Learning Targets 3 and 13: Make inferences or predictions. (2.1.4)

Stimulus Attributes:

1. Format: Literary, informational, task-oriented text

2. Content:

- For inference: A significant idea in the selection must be indirectly stated; the selection must contain sufficient clues that the idea may be inferred.
- For prediction: A selection must provide clues about a likely outcome or action without stating it.

Item Description:

1. Possible item types: M, S

2. Item format: Clearly states requirement that students use evidence from the text to grasp an important idea not directly stated in the text, or that they make a reasonable prediction of an outcome or action.

3. Response format: Response may be phrase(s), sentence(s), or paragraph(s).

4. Multiple-choice items:

Example of stem for Inference:

What conclusion can be drawn from reading this selection?

Example of response choices:

a. The correct response is an idea that can reasonably be inferred from the text.
b. An incorrect response may contain an idea that the selection does not really support.

Example of stem for Prediction:

What do you think (character) will do now that (cite circumstances at end of story)?

Example of response choices:

a. The correct response is an outcome that can reasonably be predicted given the information in the text.
b. An incorrect response is not appropriate given the information in the text.

5. Short-answer items:

Example for Inference:

Why do you think (character) did (action)? Provide evidence or examples from the story to support your answer.

How did (character) feel when (event) happened? Provide evidence or examples from the story to support your answer.

Example for Prediction:

What do you think (character) will do now that (cite circumstances at end of story)? Provide evidence or examples from the story to support your answer.

Strands LC/IC: Comprehends important ideas and details

Learning Targets 4 and 14: Interpret general and specialized vocabulary critical to the meaning of the text. (1.2.1)

Stimulus Attributes:

1. Format: Literary, informational, task-oriented text

2. Content: The selection containing the vocabulary (word, phrase, or expression) must have sufficient context clues for the reader to interpret its meaning. Note: This learning target should involve more than definition; it should involve interpretation.

Item Description:

1. Possible item types: M

2. Item format: Clearly states requirement that students interpret a word, phrase, or expression critical to the meaning of a text.

3. Response format: Word(s), phrase(s), or sentence(s).

4. Multiple-choice items:

Example of stem:

What is the meaning of (phrase or expression) in this (selection)?

Example of response choices:

a. The correct response, which may be understood from the text through thoughtful reading.
b. Incorrect responses are syntactically correct and related in some general way to ideas in the selection but that are clearly incorrect.

Example of stem:

Why does the author use the word *clever* to describe the main character in this selection?

Example of response choices:

a. The correct response requires interpretation of the author's intent.
b. Incorrect responses represent slightly incorrect interpretations of the author's intent.

Strand LA: Analyzes, interprets, and synthesizes

Learning Target 5: Apply understanding of literary elements (genres; story elements such as plot, character, setting; stylistic devices) and graphic elements/illustrations (1.4.1, 1.4.3)

Stimulus Attributes:

1. Format: Literary text
2. Content: The selection may lend itself to analysis of one or more of the following elements:
   a. Plot: A series of episodes in which a conflict is developed and resolved.
   b. Characters: The people, animals, or personified objects in the story.
   c. Setting: The location(s) and time(s) of the story.
   d. Theme: The central idea or message in a work of literature.
   e. Point of view: The vantage point from which the author presents the story.
   f. Conflict: The central problem that drives the story.
   g. Resolution: Follows climax, or the point of highest interest.
   h. Stylistic devices: For Grade 10, figurative language such as simile, metaphor, personification; exaggeration; irony; humor, sarcasm; symbols; dialogue and other ways of developing characterization; and mood.

Item Description:

1. Possible item types: M, S, E
2. Item format: Clearly states requirement that students give evidence of having analyzed how some significant literary elements interact.
3. Response format: Response may be phrase(s), sentence(s), or paragraph(s).
4. Multiple-choice items:
   Example of stem:
   How does the setting of this story impact its plot?
   Example of response choices:
   a. The correct response represents a reasonable conclusion drawn through analysis of the selection.
   b. An incorrect response may be a faulty conclusion that could be drawn through less-than-careful analysis of the selection.
5. Short-answer items:
   Example of item:
   Give an example of (literary device) from the selection. Explain the point the author makes with this (literary device).
6. Extended-response item:
   Example of item:
   Analyze how the story would be different if it had been told from (character's) point of view. Discuss two events to support your answer.
   OR
   Analyze how (character's actions or trait) contributes to the conflict in the story. Use an example from the story to support your response.

Strands LA/IA: Analyzes, interprets, and synthesizes

Learning Targets 6 and 16: Compare/contrast elements of the text or make connections within the text. (2.2.1)

Stimulus Attributes:

1. Format: Literary, informational, and task-oriented text

2. Content: The selection must contain sufficient information for the reader to identify similarities and differences in elements of the text or to make connections (such as cause and effect) within a text.

Item Description:

1. Possible item types: M, S, E

2. Item format: Clearly states requirement that students compare/contrast or make connections between and among elements within a text.

3. Response format: Responses may be phrase(s), sentence(s), or paragraph(s).

4. Multiple-choice items:

Example of stem for compare/contrast:

Which sentence tells how (two characters in the story) are similar?

Example of response choices:

a. The correct response identifies an appropriate similarity.
b. Incorrect responses do not identify similarities; they may focus on something that is true of one character or the other but not both.

Example of stem for connections within text:

Which sentence explains why (event) happened?

Example of response choices:

a. The correct response is a reasonable statement of causation.
b. Incorrect responses are events in the selection that thoughtful reading reveals are not really the cause.

5. Short-answer items:

Example of item stem for compare/contrast:

How are (two elements in the selection) similar and different?

Example of item for connections within text:

Explain why (event) happened.

6. Extended-response items:

Example of item stem for compare/contrast:

Write a description of how (characters) are similar and how they are different. Then analyze how this similarity and difference influences (a text-based outcome).

Example of item stem for connections within text:

Analyze how (events in the selection) cause (culminating event in the selection). Choose two (events) to discuss in depth.

Strands LA/IA: Analyzes, interprets, and synthesizes

Learning Targets 7 and 17: Compare/contrast or make connections between or among texts or synthesize information from a variety of resources. (2.2.1)

Stimulus Attributes:

1. Format: Literary, informational, and task-oriented text

2. Content: Groups of selections are presented to give students opportunities to make connections between or among texts; these are likely to be of the comparison/contrast variety, but may include cause/effect or other connections. These selections must contain sufficient information for the reader to analyze connections, or they must present information for students to synthesize.

Item Description:

1. Possible item types: M, S, E

2. Item format: Clearly states requirement that students make connections between or among texts (example: comparison/contrast, cause/effect) or synthesize information from a variety of resources.

3. Response format: Responses may be phrase(s), sentence(s), or paragraph(s).

4. Multiple-choice items:

Example of stem for compare/contrast between/among texts:

How does (story character's) feeling about (subject) compare to the poet's feeling about (subject)?

Example of response choices:

- The correct response identifies an appropriate similarity.

b. Incorrect responses identify elements that exist in one passage but not in the other(s)

Example of stem for cause/effect between/among texts:

After reading both the story and the article, what seems to be the reason for (event)?

Example of response choices:

a. The correct response is a reasonable statement of causation that may be gleaned from reading two or more selections carefully.
b. Incorrect responses are statements of causation that careful reading of two or more selections reveals to be unsound.

5. Short-answer items:

Example of item stem for compare/contrast between/among a poem and story:

Which character in the story is most like the speaker in the poem? Explain how they are alike and how they are different.

Write a paragraph describing how (author of Selection 1) and (author of Selection 2) would respond to (a hypothetical situation).

Example of item stem for synthesize:

How important was (item) to people in the nineteenth century? Use the story and the article to answer this question.

6. Extended-response items:

Example of item prompt for compare/contrast between/among texts:

Suppose you were to take (position on issue raised in the selections). **Discuss an important idea from each selection you would use to support your position. Explain how each idea supports your position.**

Strands LT/IT: Thinks critically

Learning Targets 8 and 18: Analyze author's purpose and evaluate effectiveness for different audiences. (2.3.2, 2.3.3)

Stimulus Attributes:

1. Format: Literary, informational, and task-oriented text

2. Content: The selection must allow consideration of author's purpose and intended audience. The selection may lend itself to analysis of one or more of the following:

a. Fact/opinion
b. Author's point of view
c. Author's tone
d. Author's craft
e. Author's use of argument and propaganda techniques (from Listening/Observing EALRs)
- Snob appeal: Appealing to social or intellectual pretensions.
- Endorsement: Basing an argument on what a famous person says.
- Name-calling: Applying a negative term or label to a person or product.
- Bandwagon: Arguing that you should do something because everyone else is doing it.
- Hasty generalization: Drawing a conclusion that is too broad or based on too little evidence.
- Poisoning the well: (same as loaded words) Using emotional words instead of arguments.
- Transfer: Trying to transfer people's good feelings about one thing to something else.
- Plain folks: The opposite of snob appeal; appealing to a desire to be plain and unpretentious.

f. Author's use of bias
g. Author's use of stereotype

NOTE: If an item calls for evaluation of the logic and reasoning of these elements in a text, the item is classified as Learning Target 9 or 19.

Item Description:

1. Possible item types: M, S, E

2. Item format: Clearly states requirement that students analyze author's purpose and/or evaluate the effectiveness of a text for different audiences. Note: In items analyzing author's use of argument and/or propaganda techniques, define terms.

3. Response format: Responses may be phrase(s), sentence(s), or paragraph(s).

4. Multiple-choice item:

Example of stem:

Which sentence in this advertisement is an example of the author's use of a "Bandwagon" argument (urging people to do something because everyone is doing it)?

Example of response choices:

a. The correct response identifies an appropriate sentence.
b. Incorrect responses identify sentences that do not represent this type of argument; they may represent other types of arguments, or they may not be representative of any type of argument at all.

5. Short-answer items:

Example of items:

Who is the audience for this poem? How do you know? Support your answer with evidence from the poem.

What is the author's purpose in this poem? How does he use (the poem's structure) to achieve his purpose? Support your answer with information from the poem.

7. Extended-response items:

Example of item:

Write (number) of the author's main arguments for her/his point of view on (issue discussed in selection). How do you think she/he would feel about (related issue)? Support your answer with evidence from the selection.

Strand LT: Thinks critically

Learning Target 9: Evaluate reasoning and ideas/themes related to the text (2.3.1, 2.3.4)

Stimulus Attributes:

1. Format: Literary text

2. Content: The selection must allow consideration and evaluation of ideas or themes related to it.

Item Description:

1. Possible item types: M, S, E

2. Item format: Clearly states requirement that students evaluate reasoning or ideas or themes related to a literary text. Conclusions and generalizations that may be drawn from the text may be the focus of this evaluation.

3. Response format: Responses may be phrase(s), sentence(s), or paragraph(s).

4. Multiple-choice items:

    Example of stem:

        What is wrong with (character's) reasoning in this story?

    Example of response choices:

        a. The correct response identifies a problem with the reasoning demonstrated by the character.
        b. Incorrect responses may identify a problem with another character's reasoning or a problem in reasoning that is not a part of the story.

5. Short-answer items:

    Example of items:

        Is the following a reasonable conclusion that may be drawn from the selection? Why or why not? Refer to the selection to support your answer. (Then state a conclusion that may or may not be reasonably drawn from the selection.)

        After reading the first part of the selection you might think (idea). Does the rest of the selection support this (idea)?

6. Extended-response items:

    Example of item:

        Analyze and evaluate the (strategies/arguments) the author uses to get her point across in this (essay). Consider what (strategies/arguments) the writer uses and how effective each (strategy/argument) is. Use evidence from the selection in your response.

Strands LT/IT: Thinks critically

Learning Targets 10 and 20: Extend information beyond text (make generalizations beyond the text to a broader idea or concept, or apply information to other texts or situations, or give a response to reading) (2.3.1, 2.3.5, 2.3.6, 2.3.7)

Note: Washington prefers focusing on the first and second parts of the target and avoiding the third.

Stimulus Attributes:

1. Format: Literary, informational, and task-oriented text

2. Content: The selection must be substantive enough to allow a student to extend information and ideas they have gotten from reading to situations or ideas beyond the text.

Item Description:

1. Possible item types: M, S, E

2. Item format: Clearly states requirement that students extend what they have learned from the reading. They may be called on to make generalizations beyond the text to a broader idea or concept or to apply information to other texts or situations, or to give a response to reading.

3. Response format: Response may be phrase(s), sentence(s) or paragraph(s).

4. Multiple-choice items:

Example of stem for generalization:

After reading this article, which of the following generalizations can you make?

Example of response choices:

a. The correct response identifies a generalization that can be logically made from reading the selection.
b. An incorrect response may identify a generalization that is too broad or based on too little evidence.
c. An incorrect response may identify a generalization that appears to be reasonable but is flawed in its logic.

5. Short-answer items:

Example of item stem for personal response to text:

Would you have wanted to travel west as pioneers did in the story? Support your answer by referring to specific events in the story.

6. Extended-response items:

Example of item stem for applying information to another text or situation:

How could you use what you learned about (subject) in this selection to (do a particular task)? Refer to specific information in the selection to explain your ideas.

How would you handle (character's) problem in the story? Support your answer by referring to specific events in the story.

Strand IC: Comprehends Important Ideas and Details

Learning Target 11: Demonstrate understanding of major ideas and supportive details. (2.1.2)

Stimulus Attributes:

1. Format: Informational and task-oriented text

2. Content: The selection must be substantive enough that major and supporting ideas can be ascertained.

Item Description:

1. Possible item types: M, S

2. Item format: Clearly states requirement that students focus on major or supporting ideas and relationship between/among them.

3. Response format: Response may be phrase(s), sentence(s), or paragraph(s).

4. Multiple-choice items:

    Example of stem:

        Which sentence gives the most important idea in the (selection)?

    Example of response choices:

        a. The correct response is the major idea in the selection.
        b. An incorrect response may be an idea that is contained in the selection but is not the major idea.
        c. An incorrect response may be an idea that is related in some way to the major idea, but it is not in the selection.

5. Short-answer items:

    Examples of items:

        Give two of the most important ideas in the selection.

Strand IA: Analyzes, interprets, and synthesizes

Learning Target 15: Apply understanding of text features (titles, headings, and other information divisions, table of contents, indexes, glossaries, prefaces, appendices, captions) and graphic features. (1.5.2)

Stimulus Attributes:

1. Format: Informational and task-oriented text
2. Content: The selection may contain one or more of the following elements:
    a. titles, headings, other information divisions
    b. table of contents
    c. index
    d. glossary
    e. preface
    f. appendix
    g. captions
    h. graphic features

Item Description:

1. Possible item types: M, S
2. Item format: Clearly states requirement that students give evidence of having analyzed some significant text feature.
3. Response format: Response may be phrase(s), sentence(s), or paragraph.
4. Multiple-choice items:

    Example of stem:

    Look at (graphic) and (locate or interpret information).

    Examples of response choices:

    a. The correct response indicates appropriate location or interpretation of information.
    b. An incorrect response may not be the best choice because it focuses on an idea not represented in the graphic.
    c. An incorrect response may be too narrow to be acceptable.
    d. An incorrect response may not be true.

5. Short-answer items:

    Example of item:

    What important information do the (photographs, captions) add to the article?

    Use the table of contents and the index to decide where to look to locate information about (topic).

Strand IT: Thinks critically

Learning Target 19: Evaluate reasoning and ideas related to the text (2.3.1, 2.3.4)

Stimulus Attributes:

1. Format: Informational and task-oriented text

2. Content: The selection must allow consideration and evaluation of ideas related to it.

Item Description:

1. Possible item types: M, S, E

2. Item format: Clearly states requirement that students evaluate reasoning or ideas related to an informational or task-oriented text. (Task-oriented texts may include display advertisements created for the purposes of the test.) Conclusions and generalizations that may be drawn from the text may be the focus of this evaluation.

3. Response format: Responses may be phrase(s), sentence(s), or paragraph(s).

4. Multiple-choice items:

    Example of stem:

        What is wrong with making the following generalization after reading the selection? (Give a generalization that is too broad to be drawn from the selection)

    Example of response choices:

        a. The correct response indicates that the generalization is too broad.

        b. Incorrect responses may identify a problem the statement does not have.

5. Short-answer items:

    Examples of items:

        Evaluate the way in which this advertisement appeals to readers. Is it an effective advertisement? Explain why or why not. Refer to specific elements of the advertisement in your explanation.

        Is the following a reasonable conclusion that may be drawn from the selection? Why or why not? Use evidence from the selection to support your answer. (Then state a conclusion that may or may not be reasonably drawn from the selection.)

        After reading the first paragraph of the selection, you might make the generalization that (state a generalization that might be reasonably made from reading the first paragraph). Does the rest of the selection support this generalization? Explain your answer with details from the selection.

6. Extended-response items:

    Example of item:

        Analyze and evaluate the (strategies) the author uses to get her point across in this (selection). Consider what (strategies) the writer uses and how effective each (strategy) is. Use evidence from the selection in your response.

APPENDIX D

General Scoring Rules for the Washington Assessment of Student Learning

Listening

Reading

Mathematics

Writing

SCORING OF OPEN-ENDED LISTENING ITEMS

Individual scoring criteria were developed for each constructed-response item. Short-answer listening items were scored on a scale of 0 to 2 points. The following scoring criteria were used to guide item writers in their development of item specific scoring criteria. This helped to ensure that the item scoring criteria were clearly focused on summarizing information and paraphrasing main ideas.

Scoring Criteria for Short Answer Listening Items

SUMMARIZING AND PARAPHRASING MAIN IDEAS:

2 A two point response shows thorough comprehension of main ideas or an accurate summary of events.

1 A one point response shows partial comprehension of main ideas or a partially accurate summary of events.

0 A 0 point response shows little or no understanding of the passage main ideas or events.

SCORING OF OPEN-ENDED READING ITEMS

Individual scoring criteria were developed for each constructed-response item. Short-answer items were scored on a scale of 0 to 2 points, and extended-response items were scored on a scale of 0 to 4 points. The following scoring criteria were used to guide item writers in their development of item specific scoring criteria. This helped to ensure that the item scoring criteria were clearly focused on the appropriate dimension of reading performance: basic comprehension of main ideas and details and analysis OR analysis, interpretation, and critical thinking about text.

Scoring Criteria for Short Answer Reading Items

MAIN IDEAS AND DETAILS:

2 A two point response:
- shows thorough comprehension of main idea and important details
- uses ample, relevant information from text(s) to support responses

1 A one point response:
- shows partial comprehension of main idea and important details (may grasp main idea but show difficulty distinguishing between important and unimportant details; may miss part of fundamental who/what/where/when/why)
- attempts to use information from text(s) to support responses; support may be limited or irrelevant

0 A 0 point response shows little or no understanding of the passage main ideas and details.

ANALYSIS, INTERPRETATION, AND CRITICAL THINKING ABOUT TEXT:

2 A two point response:
- analyzes appropriate information and/or makes thoughtful connections between whole texts/parts of texts
- develops thoughtful interpretations of text
- uses sufficient, relevant evidence from text(s) to support claims

1 A one point response:
- analyzes limited information and/or makes superficial connections between whole texts/parts of texts
- develops conventional or simplistic interpretations of text
- attempts to use evidence from text(s) to support claims; support may be limited or irrelevant

0 A 0 point response shows little or no understanding of the passage main ideas and details.

Scoring Criteria for Extended Response Reading Items

ANALYSIS, INTERPRETATION, AND THINKING ABOUT TEXT:

4 Points: Meets all relevant criteria

- thoroughly analyzes appropriate information and/or makes insightful connections between whole texts/parts of texts
- develops insightful interpretations of text
- uses ample, relevant evidence from text(s) to support claims

3 Points: Meets or most all relevant criteria

- analyzes appropriate information and/or makes thoughtful connections between whole texts/parts of texts
- develops thoughtful interpretations of text
- uses sufficient, relevant evidence from text(s) to support claims

2 Points: Meets some relevant criteria

- analyzes limited information and/or makes superficial connections between whole texts/parts of texts
- develops conventional or simplistic interpretations of text
- attempts to use evidence from text(s) to support claims; support may be limited or irrelevant

1 Point: Meets few relevant criteria

- shows difficulty analyzing information and/or makes weak connections between whole texts/parts of texts
- may not develop beyond literal interpretation of text
- uses little or no evidence to support claims

0 points - Student's response provides no evidence of interpretation or critical analysis of text required by the prompt; or the prompt may simply be recopied; or the response may be "I don't know" or a question mark (?).

SCORING OF OPEN-ENDED MATHEMATICS ITEMS

Individual scoring criteria were developed for each constructed-response item. Short-answer items were scored on a scale of 0 to 2 points, and extended-response items were scored on a scale of 0 to 4 points. The following scoring criteria were used to guide item writers in their development of item specific scoring criteria. This helped to ensure that the item scoring criteria were clearly focused on the appropriate dimension of mathematics performance: conceptual and procedural understanding, mathematical problem-solving, mathematical communication, mathematical reasoning, OR mathematical connections.

General Scoring Criteria For Short-Answer Mathematics Items

MATHEMATICAL CONCEPTS AND PROCEDURES:

2 A 2-point response shows complete understanding of the concept or task, as well as consistent and correct use of applicable information and/or procedures. Set-up and computations are accurate.

1 A 1-point response shows partial understanding of the concept or task. There may be minor errors in the use of applicable information and/or procedures. Set-up or computations may have minor errors.

0 A 0 point response shows little or no understanding of the concept or task.

COMMUNICATING MATHEMATICAL UNDERSTANDING:

2 A 2-point response shows understanding of how to effectively and appropriately interpret, organize, and/or represent mathematical information relevant to the concept.

1 A 1-point response shows some understanding of how to interpret, organize, and/or represent mathematical information relevant to the concept; however, the response is not complete or effectively presented.

0 A 0 point response shows little or no understanding of how to interpret, organize and/or represent mathematical information relevant to the concept.

SOLVING MATHEMATICAL PROBLEMS:

2 A 2-point response shows thorough investigation, clear understanding of the problem, and/or effective and viable solution.

1 A 1-point response shows partial investigation and/or understanding of the problem, and/or a partially complete or partially accurate solution.

0 A 0-point response shows very little or no investigation and/or understanding of the problem, and/or no visible solution; or the prompt may simply be recopied, or may indicate "I don't know" or a question mark (?).

General Scoring Criteria For Short-Answer Mathematics Items (Cont.)

MATHEMATICAL REASONING

2 A 2-point response shows effective reasoning through a complete analysis or thorough interpretation, supported predictions, and/or verification.

1 A 1-point response shows somewhat flawed reasoning either through incomplete analysis or interpretation, prediction that lacks support, or inadequate verification.

0 A 0-point response shows very little or no evidence of reasoning; or the prompt may simply be recopied, or may indicate "I don't know" or a question mark (?).

MAKING MATHEMATICAL CONNECTIONS:

2 A 2-point response makes clear and effective connections within and/or between conceptual or procedural areas.

1 A 1-point response makes vague or partially accurate connections within and/or between conceptual or procedural areas.

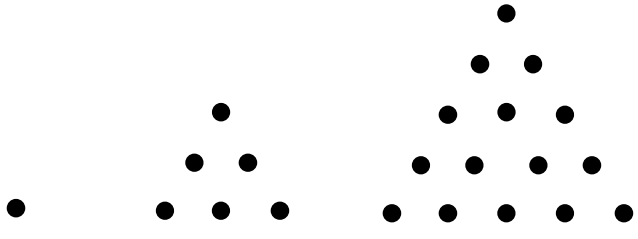
0 A 0-point response makes little or no connection within or between conceptual or procedural areas; or the prompt may simply be recopied, or may indicate "I don't know" or a question mark (?)

EXAMPLE OF SPECIFIC SCORING CRITERIA FOR A SHORT-ANSWER MATHEMATICS CONCEPTS AND PROCEDURES ITEM

Primary Essential Learning Requirement: Student ***understands and applies the concepts and procedures of mathematics: algebraic sense.***

The ancient Greeks discovered that certain numbers, when arranged in dot patterns, for definite shapes. Triangular numbers, for example, have dot patterns that can be arranged into triangles. A sequence of triangular numbers is shown below.

What is the next triangular number in this sequence? Clearly explain or show the reason for your answer.



Next triangular number: _____

SCORING CRITERIA FOR ITEM

2 A 2-point response identifies 28 as the next triangular number and provides a reasonable explanation for the choice.

1 A 1-point response identifies 28 as the next triangular number but provides no explanation OR an incomplete explanation for the choice OR identifies 21 as the next triangular number and gives an explanation for why it is a triangular number.

0 A 0 point response shows little or no understanding of number patterns or triangular numbers.

General Scoring Criteria For Extended-Response Mathematics Items

SOLVING MATHEMATICAL PROBLEMS:

4 points -- Meets all relevant criteria

- Thoroughly investigates the situation
- Uses all applicable information related to the problem
- Uses applicable mathematical concepts and procedures
- Constructs elegant, efficient, valid solution using applicable tools and workable strategies

3 points -- Meets all or most relevant criteria

- Investigates the situation
- Uses most applicable information related to the problem
- Uses applicable mathematical concepts and procedures
- Constructs viable/acceptable solution using applicable tools and workable strategies

2 points -- Meets some relevant criteria

- Investigates the situation, but may omit issues or information
- Uses some applicable information related tot he problem
- Uses some applicable mathematical concepts and procedures
- Constructs solution using applicable tools and workable strategies, solution may not completely address all issues or strategies may have flaws

1 point -- Meets few relevant criteria

- Attempts to investigate the situation
- Uses some applicable information related to the problem
- Uses few applicable mathematical concepts and procedures
- Attempts solution, however, mostly incomplete or not effective

0-points--Student's response provides no evidence of problem-solving skills or shows very little or no understanding of the task; or the prompt may simply be recopied, or the response may indicate "I don't know" or a question mark (?).

COMMUNICATING MATHEMATICAL UNDERSTANDING:

4 points -- Meets all relevant criteria

- Gathers all applicable information form appropriate sources
- Demonstrates interpretations and understandings in a clear, systematic, and organized manner
- Represents mathematical information and ideas in an effective format for the task, situation, and audience

3 points -- Meets most relevant criteria

- Gather applicable information from appropriate sources
- Demonstrates interpretations and understandings in a clear and organized manner
- Represents mathematical information and ideas in an expected format for the task, situation, and audience

2 points -- Meets some relevant criteria

- Gathers information from appropriate sources
- Demonstrates interpretation and understandings in an understandable manner
- Represents mathematical information in an acceptable format for the task, situation, and audiences

1 point -- Meets few relevant criteria

- Gathers little information from appropriate sources
- Demonstrates interpretations and understandings in a manner that may be disorganized or difficult to understand
- Represents mathematical information and ideas in a format that may be inappropriate for the task, situation, and audience.

0-points--Student's response shows little or no understanding of how to interpret, organize or represent mathematical information relevant to the concept; or the prompt may simply be recopied, or the response may indicate "I don't know" or a question mark (?).

MATHEMATICAL REASONING

4 points -- Meets all relevant criteria

- Makes insightful interpretations, comparisons, or contrasts of information form sources
- Effectively uses examples, models, facts, patterns, or relationships to validate and support reasoning.
- Makes insightful conjectures and inferences, if asked
- Systematically and successfully evaluates effectiveness of procedures and results, if asked
- Gives comprehensive support for arguments and results

3 points -- Meets most relevant criteria

- Makes thoughtful interpretations, comparisons, or contrasts of information form sources
- Uses examples, models, facts, patterns, or relationships to validate and support reasoning.
- Makes expected conjectures and inferences, if asked
- Successfully evaluates effectiveness of procedures and results, if asked
- Gives substantial support for arguments and results

2 points -- Meets some relevant criteria

- Makes routine interpretations, comparisons, or contrasts of information form sources
- Includes examples, models, facts, patterns, or relationships to validate and support reasoning.
- Conjectures and inferences, if given, may be naive
- Partially evaluates effectiveness of procedures and results, if asked
- Gives partial support for arguments and results

1 point -- Meets few relevant criteria

- Makes superficial interpretations, comparisons, or contrasts of information form sources
- Examples, models, facts, patterns, or relationships may not be included to validate and support reasoning.
- Conjectures and inferences, if given, may be naive
- Attends to wrong information and/or persists with faulty strategy when evaluating effectiveness of procedures and results
- Support for arguments and results may not be included

0-points--Student's response shows very little or no evidence of reasoning; or the prompt may simply be recopied, or the response may indicate "I don't know" or a question mark (?).

MAKING MATHEMATICAL CONNECTIONS:

4 points -- Meets all relevant criteria

- Shows a thorough understanding of links among areas of mathematics using equivalent representation AND/OR
- Identifies, analyzes, and/or applies mathematical patterns and concepts in other disciplines in a clear and insightful manner AND/OR
- Identifies, analyzes, and/or applies mathematical patterns and concepts in real-life situations in a clear and insightful manner

3 points -- Meets most relevant criteria

- Shows a general understanding of links among areas of mathematics using equivalent representation AND/OR
- Identifies, analyzes, and/or applies mathematical patterns and concepts in other disciplines in an obvious/expected manner AND/OR
- Identifies, analyzes, and/or applies mathematical patterns and concepts in real-life situations in an obvious/expected manner

2 points -- Meets some relevant criteria

- Shows a partial understanding of links among areas of mathematics using equivalent representation AND/OR
- Identifies, analyzes, and/or applies mathematical patterns and concepts in other disciplines AND/OR
- Identifies, analyzes, and/or applies mathematical patterns and concepts in real-life situations

1 point -- Meets few relevant criteria

- Shows a little understanding of links among areas of mathematics using equivalent representation AND/OR
- Identifies, mathematical patterns and concepts in other disciplines AND/OR
- Identifies applies mathematical patterns and concepts in real-life situations

0-points--Student's response makes very little or no connection within or between conceptual or procedural areas; or the prompt may simply be recopied, or the response may indicate "I don't know" or a question mark (?).

SCORING OF WRITING ITEMS

Students write in response to two prompts. Scoring criteria for two traits are applied to each piece of writing. One trait includes scoring for content, style, and organization of the writing, and the other trait includes scoring for the writing conventions (spelling, capitalization, punctuation, grammar). These scoring criteria are not adapted to the specific demands of a writing prompt since students have many choices about the topics for their writing and for the ways in which they apply stylistic elements.

CONTENT, STYLE, AND ORGANIZATION

4 points

- Maintains consistent focus on the topic and has ample supporting details
- Has logical organizational pattern and conveys a sense of wholeness and completeness
- Provides transitions which clearly serve to connect ideas
- Uses language effectively by exhibiting word choices that are engaging and appropriate for the intended audience and purpose
- Includes sentences of varied length and structure
- Allows the reader to sense the person behind the words

3 points

- Maintains adequate focus on the topic and has adequate supporting details
- Has logical organizational pattern and conveys a sense of wholeness and completeness, although some lapses may occur
- Provides adequate transitions in an attempt to connect ideas
- Uses effective language and appropriate word choices for the intended audience and purpose
- Includes sentences that are somewhat varied in length and structure
- Provides the reader with some sense of the person behind the words

2 points

- Demonstrates an awareness of the topic and includes some (or few) supporting details, but may include extraneous or loosely related material
- Shows an attempt at an organizational pattern, but exhibits little sense of wholeness and completeness
- Provides transitions that are weak and inconsistent
- Has a limited and predictable vocabulary that may not be appropriate for the intended audience and purpose
- Shows little variety in sentence length and structure
- Attempts to give the reader a sense of the person behind the words

1 point

- Presents minimal information or ideas and few supporting details which may be inconsistent or interfere with the meaning of the text
- Has little evidence of an organizational pattern or any sense of wholeness and completeness
- Provides transitions that are poorly utilized or fails to provide transitions
- Has a limited or inappropriate vocabulary for the intended audience and purpose
- Has little or no variety in sentence length and structure
- Provides the reader with little or no sense of the person behind the words

0-points--response is "I don't know"; response is a question mark (?); response is one word; response is only the title of the prompt; or the prompt is simply recopied.

SCORING OF WRITING ITEMS (Cont.)

CONTENT, STYLE, AND ORGANIZATION

2 points

- Consistently follows the rules of standard English for usage, spelling of commonly used words, capitalization, punctuation, and sentence formation
- Exhibits the use of complete and fluent sentences except where purposeful phrases or clauses are used for effect
- Indicates paragraphs consistently

1 point

- Fairly consistently follows the rules of standard English for usage, spelling of commonly used words, capitalization, punctuation, and sentence formation
- Generally exhibits the use of complete and fluent sentences except where purposeful phrases or clauses are used for effect
- Indicates paragraphs for the most part

0 points

- Basically does not follow the rules of standard English for usage, spelling of commonly used words, capitalization, punctuation, and sentence formation, although some elements may be correct
- Exhibits errors in sentence structure that impede communication
- Indicates paragraphs only to a limited degree

OR

- Response is "I don't know", a question mark, one word, only the title of the prompt, or prompt is simply recopied.

APPENDIX E

WASHINGTON ASSESSMENT OF STUDENT LEARNING

National Technical Advisory Committee

Washington State Assessment Advisory Team

National Technical Advisory Committee Members

Peter Behuniak, Director of Testing, Connecticut State Department of Education

Robert Linn, Professor, University of Colorado and UCLA/CRESST

William Mehrens, Professor, Michigan State University

Joseph Ryan, Professor, Arizona State University

Kenneth Sirotnik, Professor, University of Washington

Washington State Assessment Advisory Team

Gordon Ensign, Director of Assessment, Commission on Student Learning

Kathy Kimball, Assistant Director, Commission on Student Learning

Duncan McQuarrie, Director of Research and Evaluation, Washington State Office of the Superintendent of Public Instruction

Geoff Praeger, Director of Testing, Central Valley School District

Nancy Skerritt, Director of Curriculum, Tahoma School District

Catherine Taylor, Associate Professor, University of Washington

Joe Willhoft, Director of Research and Evaluation, Tacoma School District